2008

Modified pseudo-likelihood estimation for Markov random fields with Winsorized Poisson conditional distributions

Antonio Villanueva-Morales
Iowa State University

Modified pseudo-likelihood estimation for Markov random fields with Winsorized Poisson conditional distributions

by

Antonio Villanueva-Morales

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Statistics

Program of Study Committee:
Mark S. Kaiser, Major Professor
Max D. Morris
Petruta C. Caragea
Daniel J. Nordman
Gregory L. Tylka

Iowa State University

Ames, Iowa

2008

DEDICATION

I would like to dedicate this dissertation to the memory of my father Francisco and my mother Beatriz for their loving patience and encouragement to achieve successfully one of the most important goals in my life. I would also like to thank the National Council on Science and Technology of Mexico (CONACYT) and the Autonomous University of Chapingo for the financial assistance provided during the completion of my doctoral studies.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1. INTRODUCTION

The probability function of spatial statistical models involves, in general, an extremely awkward normalizing function of the parameters known as the partition function in statistical mechanics with the consequence that a direct approach to statistical inference through maximum likelihood (ML) is rarely possible. In order to avoid such intractability Besag (1975) introduced an alternative technique known as the method of maximum pseudo-likelihood (MPL) owing to its merit of being easy to implement. It has been mathematically demonstrated that, under suitable conditions, the MPL method yields estimators that are strongly consistent and asymptotically normally distributed around the true parameter value for large samples of various spatial processes (Geman and Graffigne 1986; Jensen and Møller 1991; Comets 1992; Jensen and Künsch 1994; Guyon 1995; Mase 1995). Maximization of Besag's pseudo-likelihood has become the method of popular choice for a wide range of applications. For example, Besag (1986) applied it to image restoration, and Goulard, Särkkä, and Grabarnik (1996) applied to marked Gibbs point processes. On the other hand, the MPL method trades away efficiency for computational ease. It has been shown that in many situations estimators produced by the MPL method are not efficient in comparison with ML estimators (Besag 1977; Geyer 1991; Geyer and Thompson 1992; Guyon and Künsch 1992; Huang and Ogata 1999, 2002). According to these studies, the MPL estimators are as good as the ML estimators in the weak interaction case, but the difference between the two becomes substantial when spatial interactions are strong. The efficiency loss can occur (Cressie 1993, p. 461) because the maximization of the pseudo-likelihood does not always yield functions of a minimal sufficient statistic, unlike for the ML method.

In two of their most recent studies Huang and Ogata (1999, 2002) address the problem of

improving the efficiency of MPL estimators while still keeping the technique computationally feasible. In the first article, Huang and Ogata (1999) studied the estimator which is produced by a single Newton-Raphson step of the Monte Carlo iteration method (see Penttinen 1984) starting from the MPL estimator. Simulation results of an Ising model and one auto-normal model on a region of a lattice showed that such estimator was close to the ML estimator in terms of its actual value, attained likelihood, and efficiency, even in the presence of strong interactions. On the other hand, that study also mentions that such an estimator may become even worse than the MPL estimator in cases where the latter takes outlying values which happen more frequently, according to the authors, in experiments with small data sets. In the second article, Huang and Ogata (2002) proposed a maximum generalized pseudo-likelihood (MGPL) method for Markov random fields on lattice. The MGPL estimator is the value of the parameter that maximizes the generalized pseudo-likelihood function (GPL). This GPL is the multivariate version of Besag's pseudo-likelihood which is constructed first by defining a group of adjacent sites for each site in the lattice and then taking the product of the multivariate conditional probability distributions (MCPD) of the groups of random variables defined on each group of adjacent sites. Simulation results from this study for an Ising and two auto-normal models showed better performance of the MGPL estimator than the MPL estimator, and the performance became better as the size of the groups of adjacent sites increased. On the other hand, it was observed that as the size of the groups of adjacent sites increased, the computing complexity for the MGPL estimator increased exponentially. The reason for this is the presence of a normalizing integral in the expression for each MCPD which has to be evaluated all over the support of the joint distribution for groups of site variables in each case. In the case of site variables with an auto-normal structure or for binary random variables like in the Ising model these normalizing integrals may often be evaluated numerically without too much effort for groups of adjacent sites of reasonable size (e.g., less than 6). However, for continuous Markov random fields other than auto-normal and discrete Markov fields with site variables assuming more than two values, an enormous effort might be required making the implementation of the MGPL method practically unfeasible even for small square lattices.

For example, in Markov random fields where each site variable, conditional on its neighbors, follows the distribution of a Winsorized Poisson random variable (Kaiser and Cressie 1997) the computation of the normalizing integrals rapidly becomes prohibitive with the size of the groups of adjacent sites and the size of the lattice, as the support of this distribution may be in the hundreds (or thousands). For instance, to implement the MGPL method with groups of adjacent sites of size five for a Markov random field with Winsorized Poisson conditional distributions with truncation parameter equal to 12 on a 20×20 square lattice we would require to evaluate 400 different normalizing integrals each expressed as the sum of 13^5 exponentials. This is clearly computationally unfeasible. Therefore, a more practical alternative to Huang and Ogata's MGPL approach is needed to improve the efficiency of Besag's MPL estimators for these situations.

In this dissertation research, we propose the method of maximum conditional pairwise pseudo-likelihood (MCPPL) for Markov random fields on lattice with Winsorized Poisson conditional distributions. The MCPPL estimators are the maximizers of the conditional pairwise pseudo-likelihood (CPPL) which is defined as the product of the bivariate conditional distributions corresponding to the cliques of size two of the Markov random field. Our MCPPL method is nearly as computationally simple as Besag's MPL method and can handle large amounts of data, and theoretical results such as consistency and asymptotic normality of the resultant estimators follow in a reasonable fashion. The MCPPL estimators are intended to be more efficient than the MPL estimators and more feasible than Huang and Ogata's MGPL estimator for Markov random fields where the latter is difficult to obtain, as in the case of the Winsorized Poisson model of Kaisser and Cressie (1997). Extensions of the MCPPL method to other spatial processes are feasible and a primary goal for future work.

The MCPPL method fits within the concept of composite likelihoods methods of Lindsay (1988). A composite likelihood is formed by adding together individual component log-likelihoods each of which is a valid marginal or conditional log-likelihood. The key utility of the composite log-likelihood is that the composite score equations form an additive estimating function that can be used to provide consistent parameter estimates in settings where a full

maximum likelihood estimator is not feasible or not available. Besag's MPL method and Huang and Ogata's MGPL method are also examples of composite likelihoods methods. Composite likelihood simply refers to the pooling of likelihood contributions in an additive fashion in circumstances where the components do not necessarily represent independent replicates. The use of composite likelihood methods allows a high dimensional likelihood to be approximated by a sum of easily evaluated lower-dimensional components.

In Chapter 2 we review some important theoretical background about a general Markov random field (MRF) and describe the spatial formulation of MRF's with Winsorized Poisson conditional distributions. Chapter 3 is intended to describe Besag's MPL method and Huang and Ogata's MGPL method, and derive specific formulae to implement them on MRF's with Winsorized Poisson conditional distributions on a region of a lattice. In Chapter 4 we introduce the MCPPL estimation method and derive its large sample properties of consistency and asymptotic normality. Simulation studies for an isotropic MRF with Winsorized Poisson conditionals are implemented to study the small sample properties of the MCPPL estimators and compare their performance with Besag's MPL estimators and Huang and Ogata's MGPL estimators. In Chapter 5 we study the performance of the MCPPL estimators in MRF's with directional dependence. Chapter 6 concludes the dissertation with some remarks.

CHAPTER 2. MARKOV RANDOM FIELD MODELS

2.1 Definition of a Markov random field

For data that occur at discrete locations in space, an assumption of mutual independence in a statistical model can be unrealistic. For temporal data, statistical dependence is often introduced by fitting a Markov model, that is, a model with the property that the 'present', conditioned on the 'past', in fact depends only on the 'immediate past'. For spatial data, an analogous notion can be formulated.

Suppose that the data are

$$\boldsymbol{Z} \equiv (Z(s_1), \ldots, Z(s_n))^T, \tag{2.1}$$

located at sites

$$\{s_i : i = 1, \ldots, n\}. \tag{2.2}$$

Assume that the data are jointly distributed according to

$$p(z(s_1), \ldots, z(s_n)) \tag{2.3}$$

from which the conditional probabilities, $p(z(s_i)|\{z(s_j) : j \neq i\})$; $i = 1, \cdots, n$, can be calculated.

The *neighborhood* N_i of the ith site is the collection of all other sites s_j $(j \neq i)$ such that

$$p(z(s_i)|\{z(s_k) : k \neq i\}) = p(z(s_i)|\{z(s_j) : j \in N_i\}); \qquad i = 1, \ldots, n. \tag{2.4}$$

In words, (2.4) says that the conditional probability of the random variable $Z(s_i)$, given values for $\{z(s_1), \ldots, z(s_{i-1}), z(s_{i+1}), \ldots, z(s_n)\}$, in fact depends only on its neighboring values $\{z(s_j) : j \in N_i\}$. In this sense, (2.4) is the spatial analogue of the temporal Markov property.

When the right-hand sides of (2.4) define (2.3), the stochastic process (2.1) is called a *Markov random field*. A very attractive feature of working with a Markov random field is that modeling can be carried out at the *local* level by specifying the neighborhoods

$$\{N_i : i = 1, \ldots, n\}, \tag{2.5}$$

and the conditional probabilities

$$p(z(s_i)|\{z(s_j) : j \in N_i\}) : i = 1, \ldots, n\}, \tag{2.6}$$

site by site. However, such specifications have to be made consistently, so that at a *global* level the joint probability (2.3) is well defined. Also, parameter estimation by maximum likelihood requires an expression for (2.3) in terms of (2.6), since parameters are specified at the local level.

The Hammersley-Clifford theorem (Hammersley and Clifford 1971; Clifford 1990) gives the form that (2.3) must take for it to be the joint probability function of a Markov random field. In Section 2.2, we state the theorem and give a sufficient condition due to Kaisser and Cressie (2000) that guarantees (2.3) can be obtained consistently from (2.6).

2.2 Hammersley-Clifford theorem and a converse result

To state the Hammersley-Clifford theorem (Hammersley and Clifford 1971; Clifford 1990) we have to define the central quantity of concern in the formulation of Markov random fields, the negpotential function. Let $\Omega \equiv \{\boldsymbol{z} : p(\boldsymbol{z}) > 0\}$ and choose $\boldsymbol{z}^* \equiv (z^*(s_1), \ldots, z^*(s_n))^T \in \Omega$ such that $p(\boldsymbol{z}^*)$ is finite. The *negpotential function* is defined as

$$Q(\boldsymbol{z}) \equiv \log\{p(\boldsymbol{z})/p(\boldsymbol{z}^*)\}; \qquad \boldsymbol{z} \in \Omega. \tag{2.7}$$

This definition of $Q(\cdot)$ is due to Kaisser and Cressie (2000). In the past (e.g., Besag 1974), $\boldsymbol{z}^*$ has been chosen to be $\mathbf{0} = (0, ..., 0)^T$, that is, $\boldsymbol{z}^* \equiv (z^*(s_1), \ldots, z^*(s_n))^T = \mathbf{0}$. Notice that the choice of $\boldsymbol{z}^*$ in (2.7) is completely general and requires only that $\boldsymbol{z}^* \in \Omega$ and $p(\boldsymbol{z}^*)$ is finite.

Note that if we can calculate $Q(\cdot)$, the joint probability function of $\boldsymbol{Z}$ is available as

$$p(\boldsymbol{z}) = \frac{\exp(Q(\boldsymbol{z}))}{\int_\Omega exp(Q(\boldsymbol{y}))\, d\mu(y)}, \tag{2.8}$$

for the appropriate measure μ (e.g., counting or Lebesgue).

Besag (1974) showed that $Q(\cdot)$ may be written as the expansion

$$\begin{aligned} Q(\boldsymbol{z}) = & \sum_{1\leq i\leq n} H_i(z(s_i)) + \sum_{1\leq i<j\leq n}\sum H_{ij}(z(s_i), z(s_j)) \\ & + \sum_{1\leq i<}\sum_{j<}\sum_{k\leq n} H_{ijk}(z(s_i), z(s_j), z(s_k)) \\ & + \cdots \\ & + H_{12\cdots n}(z(s_1), z(s_2), \ldots, z(s_n)); \qquad \boldsymbol{z} \in \Omega. \end{aligned} \tag{2.9}$$

Suppose that a value $\boldsymbol{z}^* \in \Omega$ has been chosen to define the negpotential function (2.7) on Ω. Consider the data (2.1) to be a random vector whose joint probability function satisfies the Markov random field support condition of Kaisser and Cressie (2000): let $D \equiv \{1, \ldots, n\}$, $z_{D\setminus i} \equiv (z(s_1), \ldots, z(s_{i-1}), z(s_{i+1}), \ldots, z(s_n))$, and $\Phi_i \equiv \{z_{D\setminus i} : p(z_{D\setminus i}) > 0\}$. Then the *Markov random field support condition* is simply that

$$\{z^*(s_i)\} \times \Phi_i \subseteqq \Omega; \qquad i = 1, \ldots, n. \tag{2.10}$$

From the joint probability function $p(\boldsymbol{z})$, $\boldsymbol{z} \in \Omega$, conditional probabilities can be calculated. Furthermore, and most importantly, the summands in (2.9) can be expressed in terms of conditional probabilities as long as the Markov random field support condition (2.10) is satisfied (Kaiser and Cressie 2000). For example, H-functions of first and second order can be conveniently written as

$$H_i(z(s_i)) = \log\left[\frac{p(z(s_i)|\{z^*(s_j) : j \neq i\})}{p(z^*(s_i)|\{z^*(s_j) : j \neq i\})}\right], \tag{2.11}$$

$$H_{ij}(z(s_i), z(s_j)) = \log\left[\frac{p(z(s_i)|z(s_j), \{z^*(s_k) : k \neq i, j\})p(z^*(s_i)|\{z^*(s_k) : k \neq i\})}{p(z^*(s_i)|z(s_j), \{z^*(s_k) : k \neq i, j\})p(z(s_i)|\{z^*(s_k) : k \neq i\})}\right]. \tag{2.12}$$

For specific details of these derivations and expressions for third- and higher-order H-functions, see Kaisser and Cressie (2000).

While the Markov random field support condition (2.10) is not needed for the negpotential function (2.7) to be defined (for this, one only needs a $\boldsymbol{z}^* \in \Omega$), it is needed for the validity of

the expansion (2.9), in terms of the H-functions defined by (2.11), (2.12), etc.. This condition is implied by, and hence is a weaker condition than, the frequently assumed *positivity condition* of Besag (1974), which states that $\Omega = \Omega_1 \times \cdots \times \Omega_n$, where $\Omega_i \equiv \{z_i : p(z(s_i)) > 0\}; i = 1, \ldots, n$. In words, this means that if $z(s_1), \ldots, z(s_n)$ can individually occur at sites $1, \ldots, n$, respectively, then they can occur together. Thus the positivity condition is sufficient (but not needed) to ensure the the validity of the expansion (2.9), in terms of the H-functions defined by (2.11), (2.12), etc.. It is usually satisfied in practice.

The Hammersley-Clifford theorem depends on the concept of a clique. Recall from (2.4) the definition of N_i, the neighborhood of the ith site; the elements of N_i are called *neighbors* of i. Then a *clique* is defined to be a set of sites that consists either of a single site or of sites that are all pairwise neighbors.

Theorem 1 (Hammersley and Clifford 1971)**.** Suppose that the stochastic process $\boldsymbol{Z} \equiv (Z(s_1), \ldots, Z(s_n))^T$ is a Markov random field with neighborhood structure $\{N_i : i = 1, \ldots, n\}$. Assume the Markov random field support condition (2.10). Then the H-functions, defined by (2.11), (2.12), etc., must satisfy the following property:

$$\text{if sites } i, j, \ldots, s \text{ do not form a clique, then } H_{ij\cdots s}(\cdot) \equiv 0. \tag{2.13}$$

For proof, see Theorem 2 of Kaisser and Cressie (2000); the original proof is reproduced in Clifford (1990).

The derivation of the expression (2.12) involves an arbitrary choice of the ith site or the jth site to be featured in the (conditional) probability statement. Since either could have been chosen (we actually chose the ith site in (2.12)), it is clear that the conditional probability consistency conditions must ensure that this choice is immaterial. In fact, conditional probabilities,

$$p(z(s_i)|\{z(s_j) : j \neq i\}); \qquad i = 1, \ldots, n, \tag{2.14}$$

for which H-functions are well defined through (2.11), (2.12), etc., yield a Markov random field given by (2.8) and (2.9) as the following theorem states.

Theorem 2 (Kaiser and Cressie 2000)**.** Assume that a value $\boldsymbol{z}^* \in \Omega$ for which the Markov random field support condition (2.10) is satisfied has been chosen. Then if the specified probabilities (2.14) are such that each of the H-functions, defined by (2.11), (2.12), etc., is invariant under permutations of its respective subset of variables, then the negpotential function Q defined by (2.9), defines a Markov random field, provided

$$\int_{\Omega} \exp(Q(\boldsymbol{y}))\, d\mu(y) < \infty. \tag{2.15}$$

For proof, see Theorem 3 of Kaisser and Cressie (2000).

The importance of the Hammersley-Clifford theorem (Theorem 1) to the conditional specification approach to model formulation is that it indicates precisely how many H-functions are to be used in construction of Q. In complex settings, such as image analysis applications involving hundreds or even thousands of random variables, identification of clique structure and the Hammersley-Clifford theorem make the task of constructing Q from the expansion (2.9) a manageable one (Kaiser and Cressie 2000).

By assuming that all third- and higher-order H-functions are identically 0, Besag (1974) showed how the one-parameter exponential family of conditional probabilities could be used to define a Markov random field. The Markov fields in this subclass are termed *auto-models*.

2.3 The formulation of Poisson auto-models

Let s_i denote a physical location in a geographic region of interest, and let $D \equiv \{s_i : i = 1, \ldots, n\}$ be a finite lattice (regular or irregular) defined by these sites. The random process associated with these geographic locations will be denoted as $\boldsymbol{Z} \equiv \{Z(s_i) : s_i \in D\}$. Auto-models are formulated on the basis of a Markov random field defined by the specification of a neighborhood N_i for each component of $\boldsymbol{Z}$. For discrete random variables, let the probability mass function (pmf) of $Z(s_i)$, conditional on its neighbors, be given by

$$p(z(s_i)|z(N_i)) \equiv p(z(s_i)|\{z(s_j) : s_j \in N_i\}) \tag{2.16}$$

A Poisson auto-model results from specifying that all components of $\boldsymbol{Z}$ have Poisson conditionals pmfs. As we mentioned in Section 2.2, the negpotential function $Q(\cdot)$ defined in (2.7)

is the quantity that connects such a conditionally specified model with the joint likelihood of $\boldsymbol{Z}$. If one can calculate $Q(\cdot)$, the joint pmf of $\boldsymbol{Z}$ is available through (2.8). Thus for discrete random variables we may recover the joint probability mass function $p(\boldsymbol{z})$ through

$$p(\boldsymbol{z}) = \frac{\exp(Q(\boldsymbol{z}))}{\sum_{y\in\Omega} exp(Q(\boldsymbol{y}))}. \tag{2.17}$$

In his formulation of auto-models Besag (1974) used $\boldsymbol{z}^* = \mathbf{0}$ in (2.7) to define $Q(\cdot)$. That is, define $Q(\cdot)$ as

$$Q(\boldsymbol{z}) \equiv \log\{p(\boldsymbol{z})/p(\mathbf{0})\}; \qquad \boldsymbol{z} \in \Omega. \tag{2.18}$$

From here on, the results stated for the rest of this chapter assume the definition (2.18) of $Q(\cdot)$.

Pairwise-only dependence is an assumption often made in spatial models (Besag 1974; Cressie 1993, Ch. 6); that is, write $Q(\boldsymbol{z})$ as

$$Q(\boldsymbol{z}) = \sum_{1\le i\le n} H_i(z(s_i)) + \sum_{1\le i<j\le n}\sum H_{ij}(z(s_i), z(s_j)), \tag{2.19}$$

where $H_{ij}(\cdot,\cdot)$ is zero if s_j is not an element of the set N_i. This restriction on the limits of summation is due to the Hammersley-Clifford theorem stated in Section 2.2. As we have seen, the terms in (2.19) may be written as functions of conditional pmfs through (2.10) and (2.11), respectively. Thus under the definition (2.18) of Q (i.e., $\boldsymbol{z}^* = \mathbf{0}$), the terms in (2.19) may be written, repectively, as

$$H_i(z(s_i)) = \log\left[\frac{p(z(s_i)|\{0(s_j) : j \neq i\})}{p(0(s_i)|\{0(s_j) : j \neq i\})}\right], \tag{2.20}$$

and

$$H_{ij}(z(s_i), z(s_j)) = \log\left[\frac{p(z(s_i)|z(s_j), \{0(s_k) : k \neq i, j\})p(0(s_i)|\{0(s_k) : k \neq i\})}{p(0(s_i)|z(s_j), \{0(s_k) : k \neq i, j\})p(z(s_i)|\{0(s_k) : k \neq i\})}\right], \tag{2.21}$$

where $0(s_i)$ denote the event '$Z(s_i) = 0$'.

Theorem 2 (see also the corollary to the Hammersley-Clifford theorem by Cressie 1993, p. 418) is the last general result that is needed to formulate a Poisson auto-model. Within the context of pairwise-only dependence, the result says that any specification of conditional pmfs $\{p(z(s_i)|\boldsymbol{z}(N_i)) : i = 1, \ldots, n\}$ such that the resulting terms for $H_{ij}(z(s_i), z(s_j))$ are symmetric

in i and j, leads to a unique, well-defined joint probability model for $\boldsymbol{Z}$ as long as the condition (2.15) holds, that is (for the counting measure μ), as long as

$$\sum_{\boldsymbol{y}\in\Omega} \exp(Q(\boldsymbol{y})) < \infty. \tag{2.22}$$

If this is the case, the joint pmf (and likelihood) is available through (2.17) and (2.19). From here on, we will refer to (2.22) as the *summability condition*.

The existing method to construct a Poisson auto-model depends on the following result of Besag (1974). For models in which the conditional pmfs are specified as belonging to an exponential family,

$$p(z(s_i)|z(N_i)) = \exp\{\mathcal{A}_i(z(N_i))\mathcal{B}_i(z(s_i)) - \mathcal{D}_i(z(N_i)) + \mathfrak{C}_i(z(s_i))\}, \tag{2.23}$$

then the functions $\mathcal{A}_i(\cdot)$ must satisfy

$$\mathcal{A}_i(z(N_i)) = \alpha_i + \sum_{j=1}^{n} \eta_{ij}\mathcal{B}_j(z(s_j)), \tag{2.24}$$

where $\eta_{ij} = \eta_{ji}$ for all i and j, and $\eta_{ij} = 0$ if s_j is not in the neighborhood N_i. A standard Poisson specification for the conditional pmf results from taking $\mathcal{B}_i(z(s_i)) = z(s_i)$, $\mathcal{D}_i(z(N_i)) = \exp\{\{\mathcal{A}_i(z(N_i))\}$, and $\mathfrak{C}_i(z(s_i)) = -\log(z(s_i)!)$. Using equations (2.19)-(2.21), the negpotential function for this regular Poisson model becomes

$$Q(\boldsymbol{z}) = \sum_{1\le i\le n} [\alpha_i z(s_i) - \log\{z(s_i)!\}] + \sum_{1\le i<j\le n}\sum \eta_{ij} z(s_i)z(s_j), \tag{2.25}$$

where $\eta_{ij} = 0$ if s_j is not in the neighborhood N_i. The joint support Ω is the n-fold Cartesian product of the set of nonnegative integers. Now, as a result of the summability condition (2.22) applied to $Q(\cdot)$ given by (2.25), we see that the sum in (2.22) is infinite should any one of the $\{\eta_{ij}\}$ be positive. Thus for a well-defined Poisson auto-model, we must have $\eta_{ij} \le 0$ for all i and j, which specifies that the model must contain only negative-dependence relations among the elements of $\boldsymbol{Z}$. Notice that this restriction on the signs of $\{\eta_{ij}\}$ is not needed for each individual conditional Poisson pmf to be well defined.

2.4 Models for Winsorized Poisson conditionals

Consider a random variable X with support on the nonnegative integers, and a fixed integer value $0 \leq R < \infty$. Then a truncated version of X results from defining

$$Z = X\, I(X \leq R) + R\, I(X > R), \tag{2.26}$$

where $I(\cdot)$ denotes the indicator function. The truncation described by (2.26) is called *Winsorization* and it converts X with infinite support to Z with support on the set $\{0, 1, \ldots, R\}$. This section is devoted to the review of the development of an auto-model for Winsorized Poisson random variables due to Kaisser and Cressie (1997) since, as we shall see, this development provides a valid and useful structure in which to consider positive dependencies among the elements of $\boldsymbol{Z}$. We begin by considering Winsorization for a single, regular Poisson random variable.

2.4.1 Poisson Winsorization

Consider a Poisson random variable X with pmf

$$p(x; \lambda) = \frac{\lambda^x}{x!} \exp(-\lambda),$$

for $\lambda > 0$ and $x \in \{0, 1, \ldots, \}$. Define the Winsorized random variable Z as in (2.26), for a fixed integer value $R < \infty$. Then the pmf of Z may be written as

$$p(z; \lambda, R) = \left\{ \frac{\lambda^z}{z!} \exp(-\lambda) \right\} I(z \leq R-1) + \left\{ 1 - \sum_{t=0}^{R-1} \frac{\lambda^t}{t!} \exp(-\lambda) \right\} I(z = R).$$

Now, from Taylor's formula for $\exp(\lambda)$, with what is often called the Lagrange form of the remainder (e.g., Taylor and Mann 1983, p. 100), we have that

$$\sum_{t=0}^{R-1} \frac{\lambda^t}{t!} = \exp(\lambda) - \frac{\lambda^R}{R!} \exp(\psi) \quad \text{for some}\, 0 < \psi < \lambda. \tag{2.27}$$

Hence,

$$p(z; \lambda, R) = \left\{ \frac{\lambda^z}{z!} \exp(-\lambda) \right\} I(z \leq R-1) + \left\{ \frac{\lambda^R}{R!} \exp(\psi - \lambda) \right\} I(z = R), \tag{2.28}$$

where $(\psi - \lambda) < 0$. It can be verified (see Kaiser and Cressie 1997 for details) that the expected value of Z is

$$E\{Z; \lambda, R\} = (R+1) - \sum_{z=0}^{R} \sum_{t=0}^{z} \left\{ \frac{\lambda^t}{t!} \exp(-\lambda) \right\},$$

so that $\partial E\{Z; \lambda, R\}/\partial \lambda > 0$. That is, the expected value of the Winsorized variable is strictly increasing in λ.

If R is large relative to λ, the effect of Winsorization on the first moment can be seen to be small. It can be verified (see Kaiser and Cressie 1997 for details) that the bounds for the expected value of Z are

$$\lambda - \sum_{z=R+1}^{\infty} \left\{ \frac{\lambda^{z+1}}{(z+1)!} \right\} < E\{Z; \lambda, R\} < \lambda - \sum_{z=R+1}^{\infty} \left\{ \frac{\lambda^{z+1}}{(z+1)!} \exp(-\lambda) \right\}. \tag{2.29}$$

Thus when R is large relative to λ, say $R \geq 3\lambda$, then the expected value of a Winsorized Poisson variable is near that of the regular (non-Winsorized) version and so, for modeling purposes, these expectations are nearly the same. This will remain true for the Winsorized Poisson spatial model presented in the next subsection.

2.4.2 Spatial formulation of a Winsorized Poisson auto-model

We now present the basic results from Kaiser and Cressie's formulation (Kaiser and Cressie 1997) of a spatial model for the random process $\boldsymbol{Z}$, where each component, conditional on its neighbors, follows the distribution of a Winsorized Poisson random variable. Note that in this case the support Ω in (2.22) is simply the n-fold Cartesian product of the set $\{0, 1, \ldots, R\}$.

First, notice that the Winsorized Poisson pmf (2.28) may be written in canonical exponential family form as

$$p(z; \lambda, R) = \exp\{z\theta - \mathcal{D}(\theta) - \log(z(s_i)!)\},$$

where $\theta \equiv \log(\lambda)$ and, for $0 < \psi < \exp(\theta)$,

$$\mathcal{D}(\theta) = \begin{cases} \exp(\theta) & \text{if } z \leq R-1, \\ \exp(\theta) - \psi & \text{if } z = R. \end{cases}$$

Writing the conditional pmf of each component of $\boldsymbol{Z}$ in this form gives, for $i = 1, \ldots, n$,

$$p(z(s_i)|z(N_i)) = \exp\{\mathcal{A}_i(z(N_i))z(s_i) - \mathcal{D}_i(z(N_i)) - \log(z(s_i)!)\}, \tag{2.30}$$

where

$$\mathcal{D}_i(z(N_i)) = \begin{cases} \exp\{\mathcal{A}_i(z(N_i))\} & \text{if } z(s_i) \leq R-1, \\ \exp\{\mathcal{A}_i(z(N_i))\} - \psi & \text{if } z(s_i) = R, \end{cases}$$

and $0 < \psi_i < \exp\{\mathcal{A}_i(z(N_i))\}$.

Proposition 1 (Kaiser and Cressie 1997)**.** For Winsorized Poisson conditional pmfs (2.30) where the parameter values $\{\alpha_i\}$ and $\{\eta_{ij}\}$ are not subject to any restrictions except that they be real, that $\eta_{ij} = \eta_{ji}$, and that $\eta_{ij} = 0$ if s_j is not in the neighborhood N_i, a valid model is obtained when

$$\mathcal{A}_i(z(N_i)) = \alpha_i + \sum_{s_j \in N_i} \eta_{ij} z(s_j). \tag{2.31}$$

For proof, see Kaisser and Cressie (1997).

We now obtain an expression for the negpotential function and check its summability. Substitution of (2.30) and (2.31) into (2.20) and (2.21) gives

$$H_i(z(s_i)) = \alpha_i z(s_i) - \log\{z(s_i)!\}. \tag{2.32}$$

and

$$H_{ij}(z(s_i), z(s_j)) = \eta_{ij} z(s_i) z(s_j) \tag{2.33}$$

Substitution of (2.32) and (2.33) into (2.19) yields the negpotential function

$$Q(\boldsymbol{z}) = \sum_{1 \leq i \leq n} [\alpha_i z(s_i) - \log\{z(s_i)!\}] + \sum_{1 \leq i < j \leq n} \sum \eta_{ij} z(s_i) z(s_j), \tag{2.34}$$

where $\eta_{ij} = \eta_{ji}$ and $\eta_{ij} = 0$ if s_j is not in the neighborhood of N_i. The summability condition (2.22) is easily verified for any real α_i and η_{ij} since Ω, the joint support set of $\boldsymbol{Z}$, is finite with $(R+1)^n$ elements.

Note that the negpotential function (2.34) is identical in form to that of a regular Poisson auto-model, although the conditional pmfs (2.30) used to formulate the Winsorized Poisson auto-model differ from those of the regular Poisson auto-model. The major distinction between these models is their support and hence the possibility that the dependence parameters $\{\eta_{ij}\}$ in (2.31) may be positive. It is through these parameters that neighboring random variables

impact the conditional pmf of $Z(s_i)$. Notice that $\eta_{ij} = 0$ for all i, j implies an independence model.

It was shown in Section 2.4.1 that the expected value of a Winsorized Poisson random variable with parameter λ is strictly increasing in λ. In terms of the conditional pmf (2.30), this translates to $E(Z(\boldsymbol{s_i})|\boldsymbol{Z}(N_i))$ being strictly increasing in $\exp\{\mathcal{A}_i(\cdot)\}$ or, equivalently, strictly increasing in $\mathcal{A}_i(\cdot)$. From (2.31), for positive η_{ij}, $\mathcal{A}_i(\cdot)$ is increasing in $z(s_j)$, holding fixed the values of $\{Z(s_k) : s_k \in N_i \setminus s_j\}$. Thus, for positive η_{ij}, $E(Z(s_i)|\boldsymbol{Z}(N_i))$ is increasing in $Z(s_j)$ for $s_j \in N_i$, which is one obvious way to express positive spatial dependence.

The effect of Winsorization on the conditional expected value of $Z(s_i)$ will be small if R is substantially larger than any of the exponentiated natural parameter functions $\exp\{\mathcal{A}_i(\cdot)\}$. That is, without Winsorization, $E(Z(s_i)|\boldsymbol{Z}(N_i)) = \exp\{\mathcal{A}_i(\cdot)\}$ and relation (2.29) holds with λ replaced by $\exp\{\mathcal{A}_i(\cdot)\}$.

CHAPTER 3. PSEUDO-LIKELIHOOD ESTIMATION

3.1 Maximum likelihood and the partition function

For a parameterized negpotential function $Q(\boldsymbol{z};\boldsymbol{\theta})$ the joint probability function $p_{\boldsymbol{\theta}}$ of $\boldsymbol{Z}$ is available as

$$p_\theta(\boldsymbol{z}) = C(\boldsymbol{\theta})^{-1} \exp\{Q(\boldsymbol{z};\boldsymbol{\theta})\}; \quad \boldsymbol{z} \in \Omega, \tag{3.1}$$

with the normalizing factor

$$C(\boldsymbol{\theta}) = \sum_{y\in\Omega} \exp\{Q(\boldsymbol{y};\boldsymbol{\theta})\}. \tag{3.2}$$

Thus, the *log-likelihood* of a parameter $\boldsymbol{\theta}$ is

$$l(\boldsymbol{\theta};\boldsymbol{z}) = \log p_{\boldsymbol{\theta}}(\boldsymbol{z}) = Q(\boldsymbol{z};\boldsymbol{\theta}) - \log C(\boldsymbol{\theta}). \tag{3.3}$$

Maximum likelihood estimation of $\boldsymbol{\theta}$ is generally numerically difficult due to the difficulty in evaluating the normalizing factor $C(\boldsymbol{\theta})$. This is the "partition function" of statistical mechanics, and to give an idea of its complexity it suffices to mention that a Nobel Prize of Chemistry was awarded to Lars Onsager in 1968 for just providing an approximation. The *maximum likelihood estimate* (MLE) can be computed through the use of stochastic iteration algorithms by the Monte Carlo maximum likelihood method (see Penttinen 1984; Younes 1988; Moyeed and Baddeley 1991; Geyer 1991; Geyer and Thompson 1992; Kaiser, Cressie, and Lee 2002). On the other hand, there have been few applied results using this method published in the literature. The reason (Huang and Ogata 1999) is the instability in its convergence to the solution by the iteration due to the stochastic approximation, while intensive computation is required for each step.

3.2 Estimation based on Besag's pseudo-likelihood

The *maximum pseudo-likelihood estimator* (MPLE) for spatial models was proposed by Besag (1975) as an alternative to the MLE to avoid the latter's intractability. The MPLE is the value of $\boldsymbol{\theta}$ that maximizes the *pseudo-likelihood* (PL),

$$L_{\mathcal{P}}(\boldsymbol{\theta}; \boldsymbol{z}) = \prod_{i=1}^{n} p_{\boldsymbol{\theta}}(z(s_i)|\{z(s_j) : s_j \neq s_i\})$$

defined as the direct product of conditional probabilities or conditional probability densities of the variable at each site, or equivalently its logarithm

$$l_{\mathcal{P}}(\boldsymbol{\theta}; \boldsymbol{z}) = \sum_{i=1}^{n} \log p_{\boldsymbol{\theta}}(z(s_i)|\{z(s_j) : s_j \neq s_i\}). \tag{3.4}$$

The maximum pseudo-likelihood method is much simpler to implement than the maximum likelihood method, and has become a method of popular choice. Besag (1986) and others have also argued that the MPLE may be more appropriate than the MLE because the model itself may be too simple for the data. It has been mathematically demonstrated that the MPLE is strongly consistent and asymptotically normally distributed around the true parameter for large samples (Geman and Graffigne 1986; Jensen and Møller 1991; Comets 1992; Jensen and Künsch 1994; Guyon 1995; Mase 1995; Goulard, Särkkä, and Grabarnik 1996). On the other hand, it has also been shown that the MPLE is not efficient in comparison with the MLE when the spatial interactions are strong (Besag 1977; Geyer 1991; Geyer and Thompson 1992; Guyon and Künsch 1992; Huang and Ogata 1999, 2002). Mase (1995) provides further theoretical and experimental evidence regarding the properties of the MPLE in various cases.

To improve the efficiency of the MPLE, Huang and Ogata (1999) used Markov chain Monte Carlo (MCMC) simulation to calculate the first- and second-order differentials of the likelihood functions at the MPLE value, and then considered the estimator which is obtained by the first step Newton-Raphson transformation from the MPLE. The experiment was conducted using the simulation of three different spatial models whose likelihoods are known in analytical form. The resulting estimator was closer to the MLE in terms of its actual value, attained likelihood, and efficiency, even in the presence of strong interactions. On the other hand, such study

also revealed that the proposed estimator may become even worse than the MPLE in cases where the MPLE is outlying from the true parameter. According to the authors the number of outliers of the MPLE increases for small square lattices.

3.3 Estimation based on Huang and Ogata's generalized pseudo-likelihood

Huang and Ogata (2002) generalized Besag's pseudo-ikelihood and proposed the maximum generalized pseudo-likelihood (MGPL) method for estimation in Markov random fields on lattice. In order to generalize Besag's pseudo-ikelihood, define a group $g(s_i)$ of sites adjacent to each site s_i, and let $\boldsymbol{Z}_{g(s_i)} \equiv \{Z(s_k) : k \in g(s_i)\}$ and $\boldsymbol{Z}^{g(s_i)} \equiv \{Z(s_k) : k \notin g(s_i)\}$ be the sets of random variables in and out of the adjacent sites group $g(s_i)$, respectively. The generalized pseudo-likelihood (GPL) for a realization $\boldsymbol{z}$ of the spatial process $\boldsymbol{Z}$ is defined by the product of (multivariate) conditional probabilities (or densities) of random variables $\boldsymbol{Z}_{g(s_i)}$ on the rest random variables $\boldsymbol{Z}^{g(s_i)}$, that is,

$$L_{\mathcal{G}}(\boldsymbol{\theta}; \boldsymbol{z}) = \prod_{i=1}^{n} p_{\boldsymbol{\theta}}(\boldsymbol{z}_{g(s_i)} | \boldsymbol{z}^{g(s_i)})^{1/|g(s_i)|},$$

where $|g(s_i)|$ denotes the number of sites in the set $g(s_i)$. Maximizing the GPL or its logarithm

$$l_{\mathcal{G}}(\boldsymbol{\theta}; \boldsymbol{z}) = \sum_{i=1}^{n} |g(s_i)|^{-1} \log p_{\boldsymbol{\theta}}(\boldsymbol{z}_{g(s_i)} | \boldsymbol{z}^{g(s_i)}) \tag{3.5}$$

with repect to $\boldsymbol{\theta}$ provides the maximum GPL estimator (MGPLE). In the case where $g(s_i) = \{s_i\}$, the MGPLE reduces to the MPLE of Besag. In the case where $g(s_i)$ is the set of all sites for any s_i, then the MGPLE is the MLE.

To construct the GPL or its logarithm and obtain the MGPLE for a realization $\boldsymbol{z}$ of the spatial process $\boldsymbol{Z}$ the functions $p_{\boldsymbol{\theta}}(\boldsymbol{z}_{g(s_i)} | \boldsymbol{z}^{g(s_i)})$ have to be calculated for each site s_i. Note that for a negpotential function $Q(\boldsymbol{z}; \boldsymbol{\theta})$ the formula for $p_{\boldsymbol{\theta}}(\boldsymbol{z})$ in (3.1) implies that

$$p_{\boldsymbol{\theta}}(\boldsymbol{z}_{g(s_i)} | \boldsymbol{z}^{g(s_i)}) = \frac{\exp\{Q(\boldsymbol{z}_{g(s_i)}, \boldsymbol{z}^{g(s_i)}; \boldsymbol{\theta})\}}{\sum_{t \in \Omega_{g(s_i)}} \exp\{Q(\boldsymbol{t}_{g(s_i)}, \boldsymbol{z}^{g(s_i)}; \boldsymbol{\theta})\}} \tag{3.6}$$

where $\Omega_{g(s_i)} \equiv \{t : p_{\boldsymbol{\theta}}(\boldsymbol{t}_{g(s_i)} | \boldsymbol{z}^{g(s_i)}) > 0\}$. In the case of the Markov random field model, the Hammersley-Clifford theorem implies that the conditional probability $p(\{z(s_j) : s_j \in g(s_i)\} | \text{all other site values})$ only depends upon $\boldsymbol{z}_{g(s_i)}$ and the values at sites neighboring the

sites $g(s_i)$; specifically, denote these neighboring sites as $\partial g(s_i)$. Then we have $p_{\boldsymbol{\theta}}(\boldsymbol{z}_{g(s_i)}|\boldsymbol{z}^{g(s_i)}) = p_{\boldsymbol{\theta}}(\boldsymbol{z}_{g(s_i)}|\boldsymbol{z}_{\partial g(s_i)})$.

Huang and Ogata (2002) conducted simulations using an Ising model and two auto-normal models on a square lattice to compare the performance of the MGPLE and the MPLE with respect to mean square error, efficiency, and closeness to the MLE . According to their results, the performance of the MGPLE was better than the MPLE, and the performance became better as the size of the groups of adjacent sites, $g(s_i)$, defined for each site, s_i, increased. On the other hand, as $|g(s_i)|$ increased, the computational complexity increased exponentially. The reason for this is the presence of the normalizing sum (integral in the case of pdfs), $\sum_{t\in\Omega_{g(s_i)}} \exp\{Q(\boldsymbol{t}_{g(s_i)}, \boldsymbol{z}^{g(s_i)}; \boldsymbol{\theta})\}$, in expression (3.6) for each $p_{\boldsymbol{\theta}}(\boldsymbol{z}_{g(s_i)}|\boldsymbol{z}^{g(s_i)})$. Note that the range of summation for this quantity increases exponentially with $|g(s_i)|$ and eventually becomes the partition function, $C(\boldsymbol{\theta})$, defined in (3.2) when $|g(s_i)| = n$, the set of all sites in the lattice. When $|g(s_i)|$ is small for each s_i, these functions may be evaluated without too much effort for binary valued site variables, $Z(s_i)$, like in the Ising model, or variables with auto-normal structure like the ones studied by Huang and Ogata (2002). However, this turns out not to be true in general. For example, in the case of Winsorized Poisson site variables with truncation parameter $R = 12$ each $\sum_{t\in\Omega_{g(s_i)}} \exp\{Q(\boldsymbol{t}_{g(s_i)}, \boldsymbol{z}^{g(s_i)}; \boldsymbol{\theta})\}$ contains $13^{|g(s_i)|}$ summands, so direct computation rapidly becomes prohibitive even for small lattices. For instance, to compute the GPL in this case for a 10×10 square lattice with $|g(s_i)| = 5$ for each s_i we would require the evaluation of 10^2 sums of exponentials each of 13^5 summands. In addition, note that the GPL and its derivatives would certainly require to be evaluated more than one time (each time $\boldsymbol{\theta}$ is updated) in an iterative optimization algorithm to obtain the MGPLE of the parameter $\boldsymbol{\theta}$.

3.4 Estimation based on Lindsay's composite likelihood

A composite likelihood is formed by adding together individual component log likelihoods each of which is a valid marginal or conditional log likelihood (Lindsay 1988). The feature of the problem to exploit is that the composite likelihood is *composed* of log likelihoods.

To construct a composite likelihood, one starts with a set of conditional or marginal events for which one can write log likelihoods $l_i(\boldsymbol{\theta})$, say $i = 1$ to k. The *composite log likelihood* (CL) is defined as

$$l_{cl}(\boldsymbol{\theta}) = \sum_{i=1}^{n} l_i(\boldsymbol{\theta}).$$

Let $\nabla l_{cl}(\boldsymbol{\theta})$ be the *composite score function*. To construct an estimator of the parameter $\boldsymbol{\theta}$, the composite likelihood estimator, one solves the composite score equation

$$\nabla l_{cl}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \mathcal{S}_i(\boldsymbol{\theta}) = 0,$$

where $\mathcal{S}_i = \nabla l_i(\boldsymbol{\theta})$ is the *component score*.

Since each component likelihood, l_i, is a true likelihood, it carries some of the features of ordinary likelihood. The first important feature is that the Kullback Leibler information inequality (Kullback and Leibler 1951) holds for each component log likelihood and hence for the (CL):

$$E_{\boldsymbol{\theta}_0}\{l_i(\boldsymbol{\theta})\} \leq E_{\boldsymbol{\theta}_0}\{l_i(\boldsymbol{\theta}_0)\} \Rightarrow \sup_{\boldsymbol{\theta}} E_{\boldsymbol{\theta}_0}\{l_{cl}(\boldsymbol{\theta})\} = E_{\boldsymbol{\theta}_0}\{l_{cl}(\boldsymbol{\theta}_0)\}$$

From this, under assumptions concerning the convergence of the CL function, one can show (Lindsay 1988) that maximizing the composite log likelihood leads to a consistent method of estimation. A second feature is that, under regularity conditions,

$$E_{\boldsymbol{\theta}}\{\mathcal{S}_i(\boldsymbol{\theta})\} = 0,$$

so that $\nabla l_{cl}(\boldsymbol{\theta})$ satisfies the requirement of being an *unbiased estimating function* (Lindsay 1988). Moreover, it is also in a special class characterized by having the components $\mathcal{S}_i$ to satisfy $\mathrm{var}_{\boldsymbol{\theta}}(\mathcal{S}_i) = E_{\boldsymbol{\theta}}\{-\nabla \mathcal{S}_i(\boldsymbol{\theta})\}$, and so (see Lindsay 1988) the *information* in $\nabla l_{cl}(\boldsymbol{\theta})$ satisfies:

$$\mathfrak{I}_{cl}(\boldsymbol{\theta}) = (\sum_{i=1}^{n} \mathrm{var}_{\boldsymbol{\theta}}\, \mathcal{S}_i)(\mathrm{var}_{\boldsymbol{\theta}} \sum_{i=1}^{n} \mathcal{S}_i)^{-1}(\sum_{i=1}^{n} \mathrm{var}_{\boldsymbol{\theta}}\, \mathcal{S}_i).$$

The idea here is that, under some assumptions, a Taylor series argument similar to that for maximum likelihood would show $\mathfrak{I}_{cl}^{-1}$ to be the asymptotic variance of the estimator formed by solving $\nabla l_{cl}(\boldsymbol{\theta}) = 0$.

Besag's pseudo-likelihood method (Section 3.2) and Huang and Ogata's generalized pseudo-likelihood method (Section 3.3) are key examples of successful composite likelihood approaches for parameter estimation in spatial data. Other examples include working independence generalized estimating equations for longitudinal data (Liang and Zeger 1986) and composite likelihood for binary spatial data (Heagerty and Lele 1998).

CHAPTER 4. MAXIMUM CONDITIONAL PAIRWISE PSEUDO-LIKELIHOOD ESTIMATION

4.1 Development of the conditional pairwise pseudo-likelihood method

To define the conditional pairwise pseudo-likelihood method we begin by considering a finite vector of discrete univariate random variables $\boldsymbol{Z} \equiv (Z(s_1), \ldots, Z(s_n))^T$, which are associated with sites $\{s_i : i = 1, \ldots, n\}$, respectively. Assume that a neighborhood structure $\{N_i : i = 1, \ldots, n\}$ has been defined for each site s_i, and that conditional probability mass functions $p(z(s_i)|\{z(s_j) : s_j \in N_i\}) \equiv p(z(s_i)|z(N_i))$ have been specified. For a fixed integer value $0 \leq R < \infty$, suppose that

$$p(z(s_i)|z(N_i)) = \exp\{\mathcal{A}_i(z(N_i))z(s_i) - \mathcal{D}_i(z(N_i)) - \log(z(s_i)!)\}, \tag{4.1}$$

where

$$\mathcal{D}_i(z(N_i)) = \begin{cases} \exp\{\mathcal{A}_i(z(N_i))\} & \text{if } z(s_i) \leq R-1, \\ \exp\{\mathcal{A}_i(z(N_i))\} - \psi & \text{if } z(s_i) = R, \end{cases}$$

for some $0 < \psi_i < \exp\{\mathcal{A}_i(z(N_i))\}$. That is, each component of $\boldsymbol{Z}$, conditional on its neighbors, follows the distribution of a Winsorized Poisson random variable with truncation parameter R. Under pairwise-only dependence (Kaiser and Cressie 1997), a valid model is obtained when

$$\mathcal{A}_i(z(N_i)) = \alpha_i + \sum_{s_j \in N_i} \eta_{ij} z(s_j) \tag{4.2}$$

for any real parameter values $\{\alpha_i\}$ and $\{\eta_{ij}\}$ such that $\eta_{ij} = \eta_{ji}$ and $\eta_{ij} = 0$ if $s_j \notin N_i$. The negpotential function for this conditional specified spatial model (see Section 2.4.2) may be written as

$$Q(\boldsymbol{z}) = \sum_{1 \leq i \leq n} [\alpha_i z(s_i) - \log\{z(s_i)!\}] + \sum_{\{s_i, s_j\} \in \mathcal{C}} \eta_{ij} z(s_i) z(s_j), \tag{4.3}$$

where $\mathcal{C}$ denotes the collection of all cliques of two locations, i.e., $\mathcal{C} \equiv \{\{s_i, s_j\} : s_j \in N_i\}$.

Knowledge of the negpotential function (see Section 2.3) is sufficient to allow recovery of the joint pmf function $p(\boldsymbol{z})$ for all $\boldsymbol{z} \in \Omega$ through

$$p(\boldsymbol{z}) = \frac{\exp(Q(\boldsymbol{z}))}{\sum_{y \in \Omega} exp(Q(\boldsymbol{y}))}. \tag{4.4}$$

Here Ω is the n-fold Cartesian product of the set $\{0, 1, \ldots, R\}$. From knowledge of the joint pmf of $\boldsymbol{Z}$ we can calculate the conditional pmf for any subset of variables of $\boldsymbol{Z}$ given values at the rest of the sites. Specifically, let $g \equiv \{s_i, s_j, ..., s_r\}$ be a subset of the sites $\{s_1, \ldots, s_n\}$, and define $\boldsymbol{Z}_g \equiv \{Z(s_k) : k \in g\}$ and $\boldsymbol{Z}^g \equiv \{Z(s_k) : k \notin g\}$. Then the conditional pmf of the variables $\boldsymbol{Z}_g$ given values of the variables $\boldsymbol{Z}^g$ is given by

$$p(\boldsymbol{z}_g|\boldsymbol{z}^g) = \frac{\exp\{Q(\boldsymbol{z}_g, \boldsymbol{z}^g)\}}{\sum_{t \in \Omega_g} exp\{Q(\boldsymbol{t}, \boldsymbol{z}^g)\}}, \tag{4.5}$$

where Ω_g, the support of the variables $\boldsymbol{Z}_g$, is the $|g|$-fold Cartesian product of the set $\{0, 1, \ldots, R\}$.

Let c represent any element of the set $\mathcal{C}$, the set of all cliques of two sites from the collection $\{s_1, \ldots, s_n\}$. Then the conditional pairwise pseudo-likelihood (CPPL) is defined as the product

$$L_{\mathcal{C}}(\boldsymbol{\theta}; \boldsymbol{z}) = \prod_{c \in \mathcal{C}} p_{\boldsymbol{\theta}}(\boldsymbol{z}_c|\boldsymbol{z}^c),$$

where $\boldsymbol{\theta}$ represents the vector which consists of the parameters α_i's and η_{ij}'s. Maximizing the CPPL or its logarithm

$$l_{\mathcal{C}}(\boldsymbol{\theta}; \boldsymbol{z}) = \sum_{c \in \mathcal{C}} \log p_{\boldsymbol{\theta}}(\boldsymbol{z}_c|\boldsymbol{z}^c) \tag{4.6}$$

with respect to $\boldsymbol{\theta}$ provides the maximum CPPL estimator (MCPPLE).

4.2 Large-Sample Properties of the Maximum Conditional Pairwise Pseudo-likelihood Estimator

We begin this section by noticing that the joint pmf (4.4) for the conditional specification (4.1) with parameterization (4.2) is a natural exponential family, i.e., the negpotential function (4.3) is a linear function of the parameters α_i's and η_{ij}'s. Thus it can be deduced from expression (4.5) that the conditional pmf $p(\boldsymbol{z}_g|\boldsymbol{z}^g)$ for a group of random variables $\boldsymbol{Z}_g$

associated with sites $g \equiv \{s_i, s_j, ..., s_r\} \subseteq \{s_1, \ldots, s_n\}$ is also a natural exponential family for the parameters α_i's and η_{ij}'s. This form of the joint pmf and the conditional pmfs results in many useful mathematical properties. For example, in the case when a parameterized neg-potential function $Q(\boldsymbol{z}; \boldsymbol{\theta})$ is a linear function of the parameter $\boldsymbol{\theta}$: $Q(\boldsymbol{z}; \boldsymbol{\theta}) = \boldsymbol{\theta}^T V(\boldsymbol{z})$ where $\boldsymbol{\theta}^T$ indicates the transpose of a column vector and $V(\boldsymbol{z})$ is a vector of the same dimension of $\boldsymbol{\theta}$, it may be seen that $\log p_{\boldsymbol{\theta}}(\boldsymbol{z}_g | \boldsymbol{z}^g)$ is concave in the parameters $\boldsymbol{\theta}$ (i.e., the α_i's and η_{ij}'s) by the same argument given in Guyon (1995, Sec. 5.1.1). Thus the log likelihood (3.3), the log pseudo-likelihood (3.4), the log GPL (3.5) and the log CPPL (4.5) are concave in $\boldsymbol{\theta}$, since all of them are finite sums of concave functions. This assures the uniqueness of the corresponding estimators, i.e., the MLE, MPLE, MGPLE, and MCPPLE, respectively, and the convergence of the gradient algorithms used in the associated optimization problem. In addition, these properties and the fact that the parameter space $\Theta \subseteq \Re^p$ is convex for natural exponential families (see, e.g., Lemma 2.7.1 of Lehmann and Romano 2004) are very useful tools to assess the large-sample properties of estimators of the parameter $\boldsymbol{\theta} \in \Theta$, as we will see later for the MCPPLE.

The asymptotic context for estimating the parameters of a Markov random field may arise in two different ways: (a) the data consist of a sequence of finite samples which are larger and larger pieces of an infinite sample from an infinite-volume Markov random field, (b) the data consist of a sequence of larger and larger independent, finite samples from a finite-volume Markov random field as the volume becomes larger and larger. The theoretical results we will use in this section to discuss the large-sample properties of the MCPPLE correspond to case (a). (Analogue results for case (b) can be found in Arnold and Strauss (1991); Gidas (1993) and Cox and Reid (2004).)

Following case (a) above, we now define notation and describe more specifically the framework in which we will discuss consistency and asymptotic normality of the MCPPLE for Markov random fields with Winsorized Poisson conditional distributions of the form (4.1). Let $\mathbb{Z}^2$ represent the 2-dimensional infinite spatial lattice. With each site $s_i \in \mathbb{Z}^2$, we associate a random variable $Z(s_i)$ taking values in the set $\Omega_0 = \{0, 1, \ldots, R\}$, for a fixed (known) positive

integer value R. For each parameter value $\boldsymbol{\theta} \in \Theta \subset \Re^p$ let $P_{\boldsymbol{\theta}}$ represent an infinite-volume Markov random field on $\Omega_\infty = \Omega_0^{\mathbb{Z}^2}$ which has finite range interactions (i.e., neighbors cannot be arbitrarily far apart) and a regular neighborhood structure on $\mathbb{Z}^2$. We will consider estimation of the parameter $\boldsymbol{\theta}$ from a single realization $\boldsymbol{z}$ of $P_{\boldsymbol{\theta}}$ observed in a finite "window" D of $\mathbb{Z}^2$. Let D_n denote a sequence of finite windows D_n of observation, increasing to $\mathbb{Z}^2$. For simplicity, we assume $D_n = [-n, n]^2$. Let $\boldsymbol{Z}_{D_n} \equiv \{Z(s_i) : s_i \in D_n\}$ and $\boldsymbol{Z}^{D_n} \equiv \{Z(s_i) : s_i \notin D_n\}$ the site random variables in and out of the window of observation D_n, respectively, and let $P_{n,\boldsymbol{\theta}}$ represent the conditional distribution of the variables $\boldsymbol{Z}_{D_n}$ given values of variables $\boldsymbol{Z}^{D_n}$. We assume that the infinite-volume Markov random field $P_{\boldsymbol{\theta}}$ having these specified conditional distributions $P_{n,\boldsymbol{\theta}}$ is unique for each $\boldsymbol{\theta} \in \Theta \subset \Re^p$ (absence of *phase transition*). We assume also that for $\boldsymbol{Z}^{D_n} \equiv 0$ (the case of *free boundary conditions*) the conditional distribution $P_{n,\boldsymbol{\theta}}$ of $\boldsymbol{Z}^{D_n}$ is as in (4.4) with negpotential function Q as in (4.3), where $\boldsymbol{\theta}$ consists of the parameters α_i's and η_{ij}'s. The statistical problem may be formulated as follows: we want to estimate $\boldsymbol{\theta}$ from larger and larger windows D_n and observations $\boldsymbol{Z}_{D_n}$ from a finite volume Markov random field $P_{n,\boldsymbol{\theta}}$.

4.2.1 Consistency

When the interaction between variables is translation invariant, consistency was proved in a general setup for the MLE (Gidas 1991; see also the pioneer work of Pickard 1987 and the references therein) and consistency of the MPLE for site variables with finite support (Geman and Graffigne 1987; see also Gidas 1986). Comets (1992, 1997) proved the strong consistency of a class of maximum objective estimators for exponential parametric families of Markov random fields on $\mathbb{Z}^d$, including both the MLE and MPLE, using large deviation estimates.

To show the consistency of the MCPPLE we will use the proposition following Theorem 3.4.4 of Guyon (1995): given verbatim as,

Proposition 2 (Guyon 1995)**.** Let Θ be an open convex subset of $\Re^p$, and $f_n : \Theta \mapsto \Re$ a sequence of convex functions which converge simply toward f. Assume that the minimum of f is reached only at $\boldsymbol{\theta}_0$.

Then if the minimum of f_n is reached at $\boldsymbol{\theta}_n$, $\boldsymbol{\theta}_n \to \boldsymbol{\theta}_0$.

For proof, see Guyon (1995, p. 122). Here, converge simply means uniformly.

We can use this proposition to prove consistency of the MCPPLE in the following framework. Let $P_{\boldsymbol{\theta}}$ represent a stationary and ergodic infinite-volume Markov random field over $\mathbb{Z}^2$ with local conditional distributions of the form (4.1) with parameterization (4.2). Let $\boldsymbol{z}$ be a realization of $P_{\boldsymbol{\theta}}$. Assume that we observe $\boldsymbol{z}$ on the finite square lattice $D_n = [-n, n]^2$ (i.e., we observe only a piece of $\boldsymbol{z}$). Suppose that we shrink the window over which the log CPPL function (4.6) is evaluated to $\mathring{D}_n = D_n \setminus \partial D_n$, where ∂D_n consists of those sites in D_n which have neighbors outside D_n. Let $\mathring{\mathcal{C}}_n \equiv \{\{s_i, s_j\} \in \mathring{D}_n : s_j \in N_i\}$, i.e., $\mathring{\mathcal{C}}_n$ consists of the collection of cliques of two locations from $\mathring{D}_n$. Then the log CPPL function on $\mathring{D}_n$ can be expressed as $l_{\mathring{\mathcal{C}}_n}(\boldsymbol{\theta}; \boldsymbol{z}) = \sum_{c \in \mathring{\mathcal{C}}_n} \log p_{\boldsymbol{\theta}}(\boldsymbol{z}_c | \boldsymbol{z}^c)$. Now, let $|\mathring{\mathcal{C}}_n|$ the cardinality of $\mathring{\mathcal{C}}_n$ and define

$$U_n(\boldsymbol{\theta}; \boldsymbol{z}) = -\frac{1}{|\mathring{\mathcal{C}}_n|} l_{\mathring{\mathcal{C}}_n}(\boldsymbol{\theta}; \boldsymbol{z}).$$

Thus the MCPPLE for sites in the window $\mathring{D}_n$ is the maximizer, respect to $\boldsymbol{\theta}$, of $l_{\mathring{\mathcal{C}}_n}(\boldsymbol{\theta}; \boldsymbol{z})$ or, equivalently, the minimizer of $U_n(\boldsymbol{\theta}; \boldsymbol{z})$. Let $\boldsymbol{\theta}_0$ be the true value of the parameter and define $K(\boldsymbol{\theta}_0; \boldsymbol{\theta}) = E_{\boldsymbol{\theta}_0}\left(\log \frac{p_{\boldsymbol{\theta}_0}(\boldsymbol{z}_c | \boldsymbol{z}^c)}{p_{\boldsymbol{\theta}}(\boldsymbol{z}_c | \boldsymbol{z}^c)}\right)$. Under assumptions of stationarity and ergodicity of $P_{\boldsymbol{\theta}}$, we have

$$\lim_{n \to \infty} [U_n(\boldsymbol{\theta}; \boldsymbol{z}) - U_n(\boldsymbol{\theta}_0; \boldsymbol{z})] = K(\boldsymbol{\theta}_0; \boldsymbol{\theta}) \qquad P_{\boldsymbol{\theta}_0} - a.s.$$

From here, using the Kullback and Leibler information inequality result (Kullback and Leibler 1951) we have that $K(\boldsymbol{\theta}_0; \boldsymbol{\theta}) \geq 0$ with equality if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. Now, $l_{\mathring{\mathcal{C}}_n}(\boldsymbol{\theta}; \boldsymbol{z})$ is a concave function of the parameters $\boldsymbol{\theta}$ for a finite Markov random field with negpotential function of the form (4.3). Then $U_n(\boldsymbol{\theta}; \boldsymbol{z})$ is a convex function of $\boldsymbol{\theta}$. The convergence $P_{\boldsymbol{\theta}_0}$-a.s. of the MCPPLE to $\boldsymbol{\theta}_0$ as $D_n \to \mathbb{Z}^2$ follows from proposition 2 by taking $f_n = [U_n(\boldsymbol{\theta}; \boldsymbol{z}) - U_n(\boldsymbol{\theta}_0; \boldsymbol{z})]$ and $f = K(\boldsymbol{\theta}_0; \boldsymbol{\theta})$. Note that, within a similar framework, the same argument can be applied to prove the strong consistency of the MPLE, and the MGPLE for groups of adjacent sites of the same size and shape.

4.2.2 Asymptotic Normality

First, notice that the log CPPL (4.6) is formed by adding together individual log likelihoods: the functions $\log p_{\boldsymbol{\theta}}(\boldsymbol{z}_c|\boldsymbol{z}^c),\ c \in \mathcal{C}$. Then, by definition, it is in the class of *composite log likelihood* functions studied by Lindsay (1988). For a composite log likelihood $U_n(\boldsymbol{\theta};\boldsymbol{z}))$, the *composite log likelihood estimator* of the parameter $\boldsymbol{\theta}$ is obtained by solving, for $\boldsymbol{\theta}$, the *composite score equations*

$$\nabla U_n(\boldsymbol{\theta};\boldsymbol{z})) = 0$$

The information in the composite score equations assumes the form

$$\mathcal{H}(\boldsymbol{\theta})\mathcal{J}(\boldsymbol{\theta})^{-1}\mathcal{H}(\boldsymbol{\theta}), \tag{4.7}$$

where $\mathcal{H}(\boldsymbol{\theta}) = \mathrm{E}_{\boldsymbol{\theta}}(-\nabla^2\, U_n(\boldsymbol{\theta};\boldsymbol{Z}))$ and $\mathcal{J}(\boldsymbol{\theta}) = \mathrm{Var}_{\boldsymbol{\theta}}\, \nabla U_n(\boldsymbol{\theta};\boldsymbol{Z})$. Quantity (4.7) is known as *Godambe information* or sandwich information (see Godambe 1960). Here, the information differs from the Fisher information matrix, because $\mathcal{H}(\boldsymbol{\theta}) \neq \mathcal{J}(\boldsymbol{\theta})$ as it does for maximum likelihood estimation in a regular case. This indicates loss of efficiency with respect to maximum likelihood estimation. Indeed, a consequence of the multivariate Cauchy-Schwartz inequality is that the difference between the Fisher and Godambe information matrices is a non-negative definite matrix (see Song 2007, p. 67).

Provided that the model is correctly specified, the asymptotic properties of the maximum composite likelihood estimators mainly depend on the rate at which the Godambe information converges to infinity. If this convergence is fast enough, then the maximum composite likelihood estimator will be consistent and asymptotically normally distributed (Varin 2007).

Guyon (1995, p. 123) gave quite general conditions for the asymptotic normality of estimators defined through a *contrast function*, the composite log likelihood being a special case. We restate Guyon's regularity conditions and discuss verification for the MCPPLE.

Consider the following conditions:

H1. There exists an open neighborhood $\mathcal{V}$ of $\boldsymbol{\theta}_0 \in \Re^p$ over which U_n is twice continuously differentiable, and there exists a $P_{\boldsymbol{\theta}_0}$-integrable random variable h such that for all pairs $i, j = 1, \ldots, p$ and all $\boldsymbol{\theta} \in \mathcal{V}$, $\left|\frac{\partial^2}{\partial\theta_i\partial\theta_j} U_n(\boldsymbol{\theta};\boldsymbol{Z})\right| \leq h(\boldsymbol{Z})$.

H2. There exists a sequence $a_n \to \infty$ such that $\mathcal{J}_n = \text{var}(\sqrt{a_n}\ \nabla U_n(\boldsymbol{\theta}; \boldsymbol{Z}))$ exists and satisfies

a. There exists $\mathcal{J} > \mathbf{0}$ with $\mathcal{J}_n \geq \mathcal{J}$ for all $n \geq n^*$ for some n^*.

b. $\sqrt{a_n}\ \mathcal{J}_n^{-1/2}\ \nabla U_n(\boldsymbol{\theta}; \boldsymbol{Z}) \to \mathbf{N}(\mathbf{0}, \mathbf{I}_{p\times p})$.

H3. There exists a sequence of nonstochastic $p \times p$ matrices $\mathcal{H}_n$ such that

a. There exists $\mathcal{H} > \mathbf{0}$ with $\mathcal{H}_n \geq \mathcal{H}$ for all $n \geq n^*$ for some n^*.

b. $\lim_n (\nabla^2\, U_n(\boldsymbol{\theta}_0; \boldsymbol{z}) - \mathcal{H}_n) = \mathbf{0}$ in $P_{\boldsymbol{\theta}_0}$-probability.

Given that these conditions are satisfied, we can use theorem (3.4.5) of Guyon (1995), yielding

$$\sqrt{a_n}\mathcal{J}_n^{-1/2}\mathcal{H}_n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \to \mathbf{N}(\mathbf{0}, \mathbf{I}_{p\times p}).$$

Under the assumptions of stationarity and ergodicity of the infinite-volume Markov random field $P_{\boldsymbol{\theta}}$, we can use similar arguments as the ones given for the MPLE in theorem 5.3.1 of Guyon (1995, p. 187) to prove asymptotic normality of the MCPPLE. Condition H2 requires that a central limit theorem can be applied to the sequence ∇U_n. Under the ergodicity assumption (see theorem 3.3.3 of Guyon 1995, p. 117) this is satisfied in our models by considering finite range interactions (i.e., neighbors are not arbitrarily far apart) and the fact that the functions $p_{\boldsymbol{\theta}}(\boldsymbol{z}_c|\boldsymbol{z}^c),\ c \in \mathcal{C}$, calculated from the finite-volume field $P_{n,\boldsymbol{\theta}}$ with negpotential function (4.3) are natural exponential families in the parameters $\boldsymbol{\theta}$. The latter property and the fact that the support of the site variables $Z(s_i)$ is finite ($\{1, 2, \ldots, R\}$) provide (under ergodicity) the satisfaction of conditions H1 and H3.

4.3 Simulation Results (Basic Model)

To assess the finite-sample properties of the MCPPLE and compare its performance with the MPLE and MGPLE we conducted a series of simulations for a Markov random field on a regular lattice with Winsorized Poisson conditional distributions of the form (4.1). For this study, we considered parameterization of the functions $\mathcal{A}_i(\cdot)$, $i = 1, \ldots, n$, as

$$\mathcal{A}_i(z(N_i)) = \kappa_i + \sum_{s_j \in N_i} \eta_{ij}\, \{z(s_j) - \exp(\kappa_j)\} \tag{4.8}$$

for any real parameter values $\{\kappa_i\}$ and $\{\eta_{ij}\}$ such that $\eta_{ij} = \eta_{ji}$ and $\eta_{ij} = 0$ if $s_j \notin N_i$. Kaiser (2007) calls (4.8) a *centered* parameterization for one-parameter exponential family Markov random field model. Distributing the sum and collecting constants verifies that (4.8) satisfies the structure of expression (4.2) by taking $\alpha_i = \kappa_i - \sum_{s_j \in N_i} \eta_{ij} \exp(\kappa_j)$. (See Kaiser, Caragea, and Furukawa (2004) for some consequences of centering the effects of neighbors in Markov random field models.)

It is straightforward to verify that expressions (2.20) and (2.21) for a set of conditionals of the form (4.1) with centered parameterization (4.8) give $H_i(z(s_i)) = (\kappa_i - \sum_{s_j \in N_i} \eta_{ij} \exp(\kappa_j))z(s_i) - log\{z(s_i)!\}$ and $H_{ij}(z(s_i), z(s_j)) = \eta_{ij} z(s_i) z(s_j)$, respectively. Thus the negpotential function (2.19) for this centered model may be written as

$$Q(\boldsymbol{z}) = \sum_{1 \le i \le n} \Big[\big(\kappa_i - \sum_{s_j \in N_i} \eta_{ij} \exp(\kappa_j)\big) z(s_i) - \log\{z(s_i)!\} \Big] + \sum_{\{s_i, s_j\} \in \mathcal{C}} \eta_{ij}\, z(s_i) z(s_j) \qquad (4.9)$$

where $\mathcal{C}$ denotes the collection of all cliques of two locations, i.e., $\mathcal{C} \equiv \{\{s_i, s_j\} : s_j \in N_i\}$.

To simulate the process we used the Gibbs sampler (Geman and Geman 1984). The essential idea of iterative simulation is to draw realizations of a random variable $\boldsymbol{Z}$ from a sequence of distributions that converge, as iterations continue, to the desired *target distribution* of $\boldsymbol{Z}$. The Gibbs sampler is a Markov Chain Monte Carlo (MCMC) method (Metropolis et al. 1953, Hastings 1970), it constructs (under some regularity conditions usually satisfied in practice) an ergodic Markov chain $\boldsymbol{Z}_1, \boldsymbol{Z}_2 \ldots$ having the target distribution as its equilibrium distribution. Each iteration $\boldsymbol{Z}_i$, $i \ge 1$, is generated by running over all site variables in either a systematic or a random order updating the value of just one variable at a time with a value simulated from its conditional distribution given the rest (see Geman and Geman 1984 for more details of the method). Iterative simulation has tremendous potential for summarizing awkward multivariate distributions, but it has its pitfalls; one major pitfall is what is called the "burn-in" or "warm up" problem, which refers to the question of how long to run the chain $\boldsymbol{Z}_1, \boldsymbol{Z}_2 \ldots$ on grounds that the chain may not yet have reached equilibrium (i.e., the target distribution). Gelman and Rubin (1992) propose a fully quantitative method to monitor the convergence of iterative simulation using several independent sequences, with starting points sampled from an overdispersed distribution. At each step of the iterative simulation,

they obtain, for each univariate random variable of interest, an estimate of its distribution and an estimate of the factor by which the scale of this distribution might be reduced if the simulations were continued indefinitely. This potential scale reduction is estimated by the ratio of the current variance estimate using the variance between the several sequence means to the within-sequence variance estimate. When this ratio is near 1, it is considered that the iterative simulation is close enough to convergence and that valid inference for the target distribution can be obtained using data from the next iterations.

The calculation of the MPLE, the MGPLE, and the MCPPLE was implemented through the Newton-Raphson iterative algorithm. This is an iterative method for solving nonlinear equations, such as equations whose solution determines the point at which a function takes its maximum. It begins with an initial guess for the solution. It obtains a second guess by approximating the function to be maximized in a neighborhood of the initial guess by a second-degree polynomial and then finding the location of that polynomial's maximum value. It then approximates the function in a neighborhood of the second guess by another second-degree polynomial, and the third guess is the location of its maximum. Specifically, for a smooth function $f(\boldsymbol{\theta})$, a Newton-Raphson step, starting from $\boldsymbol{\theta}$, is

$$\theta_{\text{new}} = \theta - \frac{f'(\theta)}{f''(\theta)} \qquad \text{(1-dim.)};$$

$$\boldsymbol{\theta}_{\text{new}} = \boldsymbol{\theta} - \left[\nabla^2 f(\boldsymbol{\theta})\right]^{-1} \nabla f(\boldsymbol{\theta}) \qquad (p\text{-dim.}),$$

where $\nabla f(\boldsymbol{\theta})$ is the gradient vector of the function f at $\boldsymbol{\theta}$, and $\nabla^2 f(\boldsymbol{\theta})$ is the matrix of second partial derivatives of the function f at $\boldsymbol{\theta}$, which is called a *Hessian matrix*, $\{\nabla^2 f(\boldsymbol{\theta})\}_{i,j} = \frac{\partial^2}{\partial\theta_i \partial\theta_j} f(\boldsymbol{\theta})$. In this manner, the method generates a sequence of guesses. These converge to the location of the maximum when the appropriate regularity conditions are satisfied and/or the initial guess is good. In practice, iteration proceeds until changes in $f(\cdot)$ between successive cycles are sufficiently small. In our particular case, for a simulation $\boldsymbol{z} \equiv \{z(s_1), \ldots, z(s_n)\}$ from the joint pmf $p_{\boldsymbol{\theta}}(\cdot)$, a general representation of the functions we want to maximize (i.e.,

the log PL (3.4), the log GPL (3.5), and the log CPPL (4.6)) would be

$$f(\boldsymbol{\theta};\boldsymbol{z}) = \sum_{k=1}^{m} |g_k|^{-1} \log\, p_{\boldsymbol{\theta}}(\boldsymbol{z}_{g_k}|\boldsymbol{z}^{g_k}) = \sum_{k=1}^{m} |g_k|^{-1} \log\, \frac{\exp\{Q(\boldsymbol{z}_{g_k}, \boldsymbol{z}^{g_k};\boldsymbol{\theta})\}}{\sum_{\boldsymbol{t}\in\Omega_k} \exp\{Q(\boldsymbol{t}, \boldsymbol{z}^{g_k};\boldsymbol{\theta})\}}, \tag{4.10}$$

where $g_k, k = 1,\ldots,m$, represents a group of adjacent sites of size $|g_k|$ from the collection $\{s_i : i = 1,\ldots,n\}$, $\boldsymbol{z}_{g_k} \equiv \{z(s_i) : i \in g_k\}$, $\boldsymbol{z}^{g_k} \equiv \{z(s_i) : i \notin g_k\}$, $Q(\boldsymbol{z}_{g_k}, \boldsymbol{z}^{g_k};\boldsymbol{\theta})$ is the rewritten version of $Q(\boldsymbol{z};\boldsymbol{\theta})$ for the variables located in g_k, and $\Omega_k = \Omega_0^{|g_k|}$ for $\Omega_0 \equiv \{0,\ldots,R\}$, i.e., Ω_k is the support of the conditional pmf $p_{\boldsymbol{\theta}}(\boldsymbol{z}_{g_k}|\boldsymbol{z}^{g_k})$. Then, expressions of $\frac{\partial}{\partial\theta_i} f(\boldsymbol{\theta};\boldsymbol{z})$ and $\frac{\partial^2}{\partial\theta_i\partial\theta_j} f(\boldsymbol{\theta};\boldsymbol{z})$ that need to be calculated at each iteration step of the Newton-Raphson algorithm for $f(\boldsymbol{\theta};\boldsymbol{z})$ as in (4.10) are, respectively,

$$\frac{\partial}{\partial\theta_i} f(\boldsymbol{\theta};\boldsymbol{z}) = \left(\sum_{k=1}^{m} |g_k|^{-1} \frac{\partial}{\partial\theta_i} Q(\boldsymbol{z}_{g_k}, \boldsymbol{z}^{g_k};\boldsymbol{\theta})\right) - \sum_{k=1}^{m} |g_k|^{-1} \sum_{\boldsymbol{t}\in\Omega_k} \frac{\partial}{\partial\theta_i} Q(\boldsymbol{t}, \boldsymbol{z}^{g_k};\boldsymbol{\theta}) \frac{\exp\{Q(\boldsymbol{t}, \boldsymbol{z}^{g_k};\boldsymbol{\theta})\}}{\sum_{\boldsymbol{t}\in\Omega_k} \exp\{Q(\boldsymbol{t}, \boldsymbol{z}^{g_k};\boldsymbol{\theta})\}} \tag{4.11}$$

and

$$\begin{aligned}\frac{\partial^2}{\partial\theta_i\partial\theta_j} f(\boldsymbol{\theta};\boldsymbol{z}) = {} & \left(\sum_{k=1}^{m} |g_k|^{-1} \frac{\partial^2}{\partial\theta_i\partial\theta_j} Q(\boldsymbol{z}_{g_k}, \boldsymbol{z}^{g_k};\boldsymbol{\theta})\right) \\ & - \sum_{k=1}^{m} |g_k|^{-1} \left\{ \sum_{\boldsymbol{t}\in\Omega_k} \frac{\partial^2}{\partial\theta_i\partial\theta_j} Q(\boldsymbol{t}, \boldsymbol{z}^{g_k};\boldsymbol{\theta}) \frac{\exp\{Q(\boldsymbol{t}, \boldsymbol{z}^{g_k};\boldsymbol{\theta})\}}{\sum_{\boldsymbol{t}\in\Omega_k} \exp\{Q(\boldsymbol{t}, \boldsymbol{z}^{g_k};\boldsymbol{\theta})\}} \right. \\ & + \sum_{\boldsymbol{t}\in\Omega_k} \frac{\partial}{\partial\theta_i} Q(\boldsymbol{t}, \boldsymbol{z}^{g_k};\boldsymbol{\theta}) \frac{\partial}{\partial\theta_j} Q(\boldsymbol{t}, \boldsymbol{z}^{g_k};\boldsymbol{\theta}) \frac{\exp\{Q(\boldsymbol{t}, \boldsymbol{z}^{g_k};\boldsymbol{\theta})\}}{\sum_{\boldsymbol{t}\in\Omega_k} \exp\{Q(\boldsymbol{t}, \boldsymbol{z}^{g_k};\boldsymbol{\theta})\}} \\ & \quad - \left(\sum_{\boldsymbol{t}\in\Omega_k} \frac{\partial}{\partial\theta_i} Q(\boldsymbol{t}, \boldsymbol{z}^{g_k};\boldsymbol{\theta}) \frac{\exp\{Q(\boldsymbol{t}, \boldsymbol{z}^{g_k};\boldsymbol{\theta})\}}{\sum_{\boldsymbol{t}\in\Omega_k} \exp\{Q(\boldsymbol{t}, \boldsymbol{z}^{g_k};\boldsymbol{\theta})\}} \right) \\ & \quad \left. \times \left(\sum_{\boldsymbol{t}\in\Omega_k} \frac{\partial}{\partial\theta_j} Q(\boldsymbol{t}, \boldsymbol{z}^{g_k};\boldsymbol{\theta}) \frac{\exp\{Q(\boldsymbol{t}, \boldsymbol{z}^{g_k};\boldsymbol{\theta})\}}{\sum_{\boldsymbol{t}\in\Omega_k} \exp\{Q(\boldsymbol{t}, \boldsymbol{z}^{g_k};\boldsymbol{\theta})\}} \right) \right\}. \end{aligned} \tag{4.12}$$

We implemented two simulation experiments to study the finite-sample properties of the MCPPLE and compare its performance with the MPLE and MGPLE under two different simplifications of the spatial dependence structure of model (4.9) on the regular square lattice $\mathbb{Z}^2_{\mathcal{L}\times\mathcal{L}} \equiv \{(u_i, v_i) : (u_i, v_i) \in \{1, 2, \ldots, \mathcal{L}\} \times \{1, 2, \ldots, \mathcal{L}\}\}$ with a four-nearest neighbor configuration for the neighborhoods N_i, $i = 1,\ldots,n$; $n \equiv |\mathbb{Z}^2_{\mathcal{L}\times\mathcal{L}}|$. We assume free boundary conditions

for the square lattice; that is, $Z((u_i, v_i)) = 0$ whenever $(u_i, v_i) \notin \mathbb{Z}^2_{\mathcal{L}\times\mathcal{L}}$. Let $s_i \equiv (u_i, v_i)$ were u_i denotes the integer value for horizontal grid location and v_i the corresponding integer for vertical grid location. In the first simulation study, we considered an isotropic (unidirectional dependence) model by modeling the set of dependencies $\{\eta_{ij} : i, j = 1, \ldots, n\}$ as

$$\eta_{ij} = \begin{cases} \eta & \text{if } s_j = (u_i - 1, v_i) \\ \eta & \text{if } s_j = (u_i + 1, v_i) \\ \eta & \text{if } s_j = (u_i, v_i - 1) \\ \eta & \text{if } s_j = (u_i, v_i + 1) \\ 0 & \text{otherwise,} \end{cases} \tag{4.13}$$

i.e., $\eta_{ij} = \eta$ whenever site s_j is in the neighborhood N_i of site s_i, and $\eta_{ij} = 0$ otherwise. In addition, we made $\kappa_i = \kappa, i = 1, \ldots, n$. Then the negpotential function (4.9) becomes

$$Q(\boldsymbol{z}) = \sum_{1\le i\le n} \Big[\big(\kappa - \eta \exp(\kappa)|N_i|\big) z(s_i) - \log\{z(s_i)!\} \Big] + \eta \sum_{\{s_i, s_j\}\in\mathcal{C}} z(s_i) z(s_j), \tag{4.14}$$

where $|N_i|$ denotes the cardinality of N_i, i.e., the number of neighbors of site s_i, $i = 1, \ldots, n$. Note that for the situation under consideration, $|N_i| = 4$ if s_i is an internal site, and $|N_i| = 3$ or 2 if s_i is at the boundary of the system (i.e., the lattice edges). In the second simulation study, we considered a model with directional dependence structure which will be described more fully in Chapter 5.

In the following, we discuss results from the first simulation study. We used the Gibbs sampler (Geman and Geman 1984), roughly described before, to generate the process induced by the negpotential function (4.14) defined over the 10×10 square lattice $\mathbb{Z}^2_{10\times 10} \equiv \{(u_i, v_i) : (u_i, v_i) \in \{1, 2, \ldots, 10\} \times \{1, 2, \ldots, 10\}\}$. In our simulation, the parameter η assumes values $0.01, 0.02, 0.03, 0.04$, and 0.05, while the true κ value is set to be always $\log 3$. Both η and κ are assumed unknown when these are estimated. The truncation parameter R of the local conditional distributions $p(z(s_i)|z(N_i))$ in (4.1)) was set fixed and known to $R = 12$ in all cases. Convergence of the iterative simulation was monitored using the method of multiple sequences proposed by Gelman and Rubin (1992). Convergence to the target distribution (i.e., the joint

pmf induced by (4.14)) was fast for all parameter values and the burn-in period for the chain was set at 400 iterations in all cases. After the burn-in period, the chain was subsampled every 10 iterations to reduce autocorrelation. For each parameter η, we generated $10,000$ realizations of the process, and from each of these we calculated the MPLE, the MGPLE, and the MCPPLE of the parameter $\boldsymbol{\theta} \equiv (\kappa, \eta)^T$ (i.e., we obtained 10,000 realizations of each estimator). The MGPLE was calculated from a GPL constructed for groups of sites of the form $g_i = \{(u_i, v_i), (u_i, v_i \pm 1), (u_i \pm 1, v_i)\}$, $i = 1, \ldots, n$; i.e., g_i is formed by the site $s_i \equiv (u_i, v_i)$ and its four-nearest neighbors (where applicable). Hereafter, $\hat{\eta}_{\mathcal{P}}$, $\hat{\eta}_{\mathcal{G}}$, and $\hat{\eta}_{\mathcal{C}}$, denote, respectively, the MPLE, MGPLE, and the MCPPLE of the parameter η, and $\hat{\kappa}_{\mathcal{P}}$, $\hat{\kappa}_{\mathcal{G}}$, and $\hat{\kappa}_{\mathcal{C}}$, the corresponding estimates for the parameter κ.

In order to compare the performance of the MPLE, MGPLE, and MCPPLE respect to precision and accuracy, we calculated estimates of the *mean square error*: $E_{\boldsymbol{\theta_0}}(\hat{\theta} - \theta_0)^2$, *bias*: $E_{\boldsymbol{\theta_0}}(\hat{\theta} - \theta_0)$, and *standard error*: $\sqrt{E_{\boldsymbol{\theta}_0}(\hat{\theta} - E_{\boldsymbol{\theta}_0}(\hat{\theta}))^2}$; where, $\boldsymbol{\theta}_0 \equiv (\kappa_0, \eta_0)^T$ represents the true value of the parameter vector $\boldsymbol{\theta} \equiv (\kappa, \eta)^T$, $E_{\boldsymbol{\theta_0}}$ denotes expectation under $\boldsymbol{\theta}_0$, the scalar θ_0 denotes one of κ_0 and η_0, and $\hat{\theta}$ represents the estimator of θ_0. For $M = 10,000$, these were estimated, respectively, by $\sum_{i=1}^{M}(\hat{\theta}_i - \theta_0)^2/M$, $\sum_{i=1}^{M}(\hat{\theta}_i - \theta_0)/M$, and $\sqrt{\sum_{i=1}^{M}(\hat{\theta}_i - \bar{\hat{\theta}})^2/(M-1)}$; $\bar{\hat{\theta}} = \sum_{i=1}^{M}\hat{\theta}_i/M$. We also computed approximate confidence intervals for the mean square error of the estimator $\hat{\theta}$ by using the central limit theorem, taking $\sqrt{\sum_{j=1}^{M}[(\hat{\theta}_j - \theta_0)^2 - \sum_{i=1}^{M}(\hat{\theta}_i - \theta_0)^2/M]^2/(M-1)}/\sqrt{M}$ as the standard error of the estimator $\sum_{i=1}^{M}(\hat{\theta}_i - \theta_0)^2/M$.

Table 4.1 gives the estimates of the mean square errors of the three estimators of the parameter κ, and Figure 4.1 shows the results graphically. We can clearly see from Figure 4.1 that the MPLE of κ ($\hat{\kappa}_{\mathcal{P}}$) has the largest values of mean square error among the three estimators, for all values considered for the parameter η. Also note that the difference of the mean square error values for $\hat{\kappa}_{\mathcal{P}}$ and the corresponding values for the mean square error of the MGPLE ($\hat{\kappa}_{\mathcal{G}}$) and the MCPPLE ($\hat{\kappa}_{\mathcal{C}}$) becomes larger as the dependence parameter η goes from 0.01 to 0.05, with the largest difference for the value $\eta_0 = 0.05$. On the other hand, we can also see from Figure 4.1 that the mean square error curves for $\hat{\kappa}_{\mathcal{G}}$ and $\hat{\kappa}_{\mathcal{C}}$, appear undistinguishable

Table 4.1 Monte Carlo estimates of *mean square errors* ($\times 10^2$) for estimators of the parameter κ in the isotropic Winsorized Poisson model

κ_0	η_0	$\hat{\kappa}_{\mathcal{P}}$	$\hat{\kappa}_{\mathcal{C}}$	$\hat{\kappa}_{\mathcal{G}}$
$\log 3$	0.01	0.369 (0.359, 0.378)	0.357 (0.349, 0.365)	0.356 (0.348, 0.363)
$\log 3$	0.02	0.474 (0.449, 0.499)	0.423 (0.412, 0.435)	0.424 (0.412, 0.436)
$\log 3$	0.03	0.534 (0.517, 0.552)	0.484 (0.473, 0.495)	0.486 (0.473, 0.499)
$\log 3$	0.04	0.715 (0.686, 0.745)	0.591 (0.576, 0.607)	0.593 (0.577, 0.609)
$\log 3$	0.05	1.107 (1.057, 1.157)	0.847 (0.825, 0.868)	0.811 (0.792, 0.830)

NOTE: Values in parentheses are the bounds of an approximate 90% confidence interval.

from each other when η_0 goes from 0.01 to 0.04, and only at the value $\eta_0 = 0.05$ does the mean square error of $\hat{\kappa}_{\mathcal{G}}$ appear to be smaller than the one for $\hat{\kappa}_{\mathcal{C}}$. Table 4.1 gives approximate 90% confidence intervals for the Monte Carlo estimates of mean square errors for the three estimators of the parameter κ. Note that values within the range covered by the interval for the mean square error of $\hat{\kappa}_{\mathcal{P}}$ are larger than the values covered by the intervals for the mean square errors of the two other estimators, with overlapping occurring only at $\eta_0 = 0.01$. Note also that the intervals for the mean square error of $\hat{\kappa}_{\mathcal{C}}$ overlap almost entirely with the corresponding intervals for the mean square error of $\hat{\kappa}_{\mathcal{G}}$, for all values considered for the dependence parameter η.

Estimates of the mean square error of the three estimators of the parameter η are given in Table 4.2 for dependence parameter values $\eta_0 = 0.01, 0.02, 0.03, 0.04$, and 0.05, and Figure 4.2 shows the results graphically. As in the case of estimators of κ, we can see clearly from Figure 4.2 that the mean square error of the MPLE of η ($\hat{\eta}_{\mathcal{P}}$) is larger than the one for the MCPPLE ($\hat{\eta}_{\mathcal{C}}$) and the one for the MGPLE ($\hat{\eta}_{\mathcal{G}}$), for each of the η_0 values. However, note that in contrast to what was observed for estimators of the parameter κ, the difference in mean square error values of $\hat{\eta}_{\mathcal{P}}$ with the corresponding values for $\hat{\eta}_{\mathcal{C}}$ and $\hat{\kappa}_{\mathcal{G}}$ appears to decrease as η_0 goes from 0.01 to 0.05, with the largest difference at $\eta_0 = 0.01$ and the smallest at $\eta_0 = 0.05$. Figure 4.2 also shows that the mean square error values for $\hat{\eta}_{\mathcal{G}}$ were the smallest of the three estimators

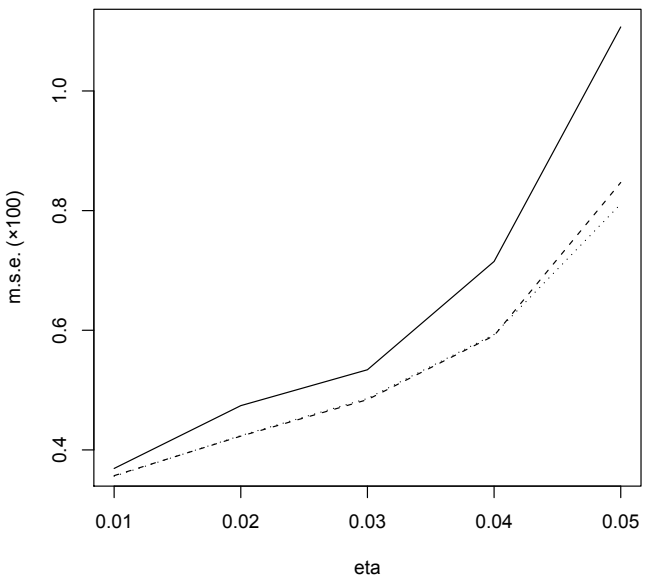

Figure 4.1 Monte Carlo estimates of *mean square errors* (m.s.e.) for estimators of the parameter κ in the isotropic Winsorized Poisson model. The solid curve corresponds to Besag's MPLE ($\hat{\kappa}_{\mathcal{P}}$), the dotted curve corresponds to Huang and Ogata's MGPLE ($\hat{\kappa}_{\mathcal{G}}$), and the dashed curve corresponds to the MCPPLE ($\hat{\kappa}_{\mathcal{C}}$).

Table 4.2 Monte Carlo estimates of *mean square errors* ($\times 10^2$) for estimators of the parameter η in the isotropic Winsorized Poisson model

κ_0	η_0	$\hat{\eta}_{\mathcal{P}}$	$\hat{\eta}_{\mathcal{C}}$	$\hat{\eta}_{\mathcal{G}}$
$\log 3$	0.01	0.065 (0.063, 0.066)	0.059 (0.058, 0.061)	0.057 (0.055, 0.058)
$\log 3$	0.02	0.062 (0.060, 0.063)	0.057 (0.056, 0.058)	0.055 (0.053, 0.056)
$\log 3$	0.03	0.057 (0.055, 0.058)	0.053 (0.052, 0.055)	0.051 (0.050, 0.052)
$\log 3$	0.04	0.050 (0.048, 0.051)	0.047 (0.046, 0.048)	0.045 (0.044, 0.046)
$\log 3$	0.05	0.039 (0.038, 0.041)	0.038 (0.037, 0.039)	0.036 (0.035, 0.037)

NOTE: Values in parentheses are the bounds of an approximate 90% confidence interval.

for all values of η_0. Note that for values of $\eta_0 = 0.01, 0.02, 0.03$, and 0.04, the mean square error values for $\hat{\eta}_{\mathcal{C}}$ keep closer, and roughly at the same distance for all η_0, to the corresponding values of mean square error of $\hat{\eta}_{\mathcal{G}}$ than to the values of mean square error of $(\hat{\eta}_{\mathcal{P}})$. Table 4.2 gives also approximate confidence intervals for the mean square error of the three estimators of η. Note that all values in the range covered by the interval for the mean square error of $(\hat{\eta}_{\mathcal{P}})$ are smaller than any value in the ranges covered by the corresponding intervals for the mean square error of $\hat{\eta}_{\mathcal{C}}$ and $\hat{\eta}_{\mathcal{G}}$ for the values $\eta_0 = 0.01, 0.02, 0.03$, and 0.04, with a slight overlapping at $\eta_0 = 0.05$ with the interval for the mean square error of $\hat{\eta}_{\mathcal{C}}$. Note also that there is not overlapping of intervals for the mean square error between $\hat{\eta}_{\mathcal{C}}$ and $\hat{\eta}_{\mathcal{G}}$ for all values η_0, with the intervals for the mean square error of $\hat{\eta}_{\mathcal{C}}$ consisting of slightly larger values.

Tables 4.3 and 4.5 give, respectively, estimates of standard errors and biases for estimators of the parameter κ, and Figures 4.3 and 4.5 show, respectively, the same results graphically. The corresponding results for estimators of the parameter η are presented in Tables 4.4 and 4.6, and Figures 4.4 and 4.6, respectively. We can see from Figures 4.3 and 4.4 a similar pattern for the standard errors of the three estimators of κ and η as what was observed for the mean square errors. The MPLEs $(\hat{\kappa}_{\mathcal{P}}, \hat{\eta}_{\mathcal{P}})$ and the MGPLEs $(\hat{\kappa}_{\mathcal{G}}, \hat{\eta}_{\mathcal{G}})$ have, respectively, the largest and the smallest standard errors for all the values considered for η_0, with the MCPPLEs $(\hat{\kappa}_{\mathcal{C}}, \hat{\eta}_{\mathcal{C}})$ having significantly better performance than the MPLEs and close to the MGPLEs

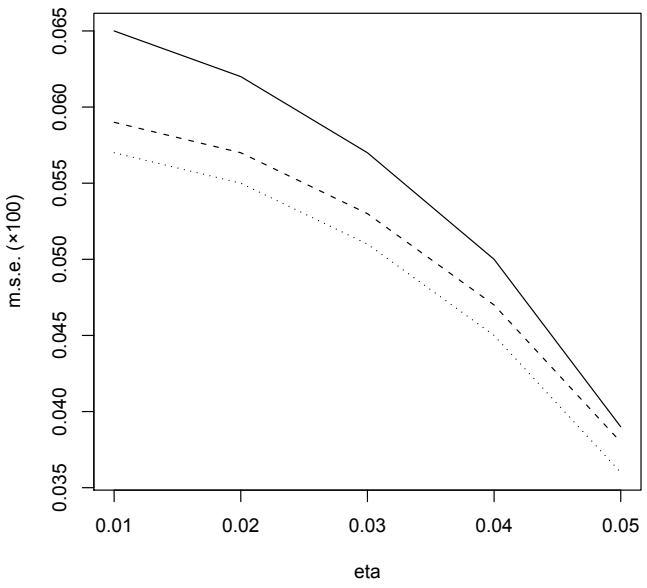

Figure 4.2 Monte Carlo estimates of *mean square errors* (m.s.e.) for estimators of the parameter η in the isotropic Winsorized Poisson model. The solid curve corresponds to Besag's MPLE ($\hat{\eta}_{\mathcal{P}}$), the dotted curve corresponds to Huang and Ogata's MGPLE ($\hat{\eta}_{\mathcal{G}}$), and the dashed curve corresponds to the MCPPLE ($\hat{\eta}_{\mathcal{C}}$).

Table 4.3 Monte Carlo estimates of *standard errors* ($\times 10^2$) for estimators of the parameter κ in the isotropic Winsorized Poisson model

κ_0	η_0	$\hat{\kappa}_{\mathcal{P}}$	$\hat{\kappa}_{\mathcal{C}}$	$\hat{\kappa}_{\mathcal{G}}$
$\log 3$	0.01	6.050	5.957	5.944
$\log 3$	0.02	6.869	6.485	6.493
$\log 3$	0.03	7.289	6.939	6.958
$\log 3$	0.04	8.433	7.672	7.686
$\log 3$	0.05	10.506	9.191	9.000

in all cases. On the other hand, we can see from Figures 4.5 and 4.6, respectively, that the pattern for bias appears to be different for the estimators of κ and η. From Figure 4.5 we can see that for estimators of κ, the best performance is for $\hat{\kappa}_{\mathcal{G}}$ with the smallest bias of the three estimators, whereas for estimators of η, it is the worst behaved with the largest bias for all values η_0 as shown by Figure 4.6. Also, Figure 4.5 shows that for estimators of κ, $\hat{\kappa}_{\mathcal{P}}$ is the worst behaved with the largest bias of the three estimators, whereas for estimators of η, it has the best performance with the smallest bias for all values η_0 as we can see from Figure 4.6. Note that for estimators of κ (Figure 4.5), the performance of $\hat{\kappa}_{\mathcal{C}}$ is roughly in the middle with values of bias in the middle of the values for $\hat{\kappa}_{\mathcal{G}}$ and $\hat{\kappa}_{\mathcal{P}}$, whereas for estimators of η (Figure 4.6), its performance is worse than the one for $\hat{\kappa}_{\mathcal{P}}$ and slightly better than the one for $\hat{\kappa}_{\mathcal{G}}$. Note also that, either for estimators of κ (Figure 4.5) or estimators of η (Figure 4.6), as η_0 goes from 0.01 to 0.05 the differences in bias between the three estimators appear to increase, with the largest effect taking place from $\eta_0 = 0.03$ to $\eta_0 = 0.05$.

A prediction study was also conducted to compare the performance of the MCPPLE to the MPLE and MGPLE. For each of the $M = 10,000$ realizations $\boldsymbol{z} \equiv (z(s_1), \ldots, z(s_{100}))$ generated under each value of the parameter $\boldsymbol{\theta} = (\kappa, \eta)^T$, we computed the quantity

$$\frac{1}{100} \sum_{i=1}^{100} \left[z(s_i) - \hat{E}_{\boldsymbol{\theta}}(Z(s_i)|\boldsymbol{z}(N_i)) \right]^2, \tag{4.15}$$

where

$$\hat{E}_{\boldsymbol{\theta}}(Z(s_i)|\boldsymbol{z}(N_i)) = (R+1) - \sum_{y=0}^{R} \sum_{t=0}^{y} \left(\frac{\hat{\lambda}_i^t}{t!} \exp(-\hat{\lambda}_i) \right) \tag{4.16}$$

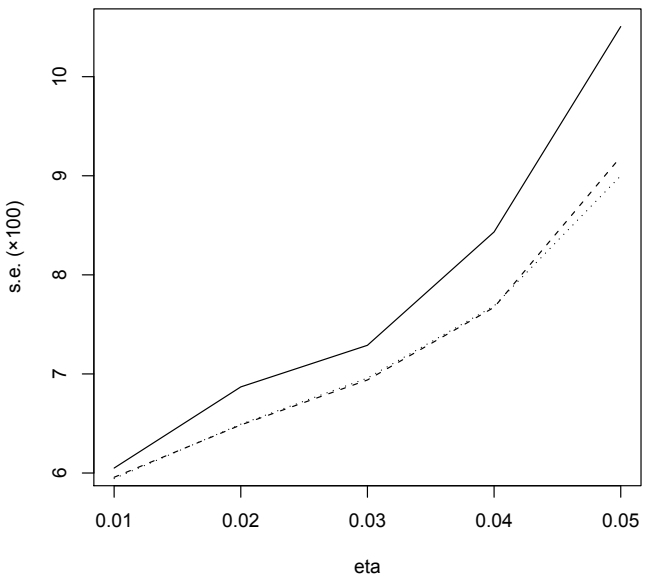

Figure 4.3 Monte Carlo estimates of *standard errors* (s.e.) for estimators of the parameter κ in the isotropic Winsorized Poisson model. The solid curve corresponds to Besag's MPLE ($\hat{\kappa}_{\mathcal{P}}$), the dotted curve corresponds to Huang and Ogata's MGPLE ($\hat{\kappa}_{\mathcal{G}}$), and the dashed curve corresponds to the MCPPLE ($\hat{\kappa}_{\mathcal{C}}$).

Table 4.4 Monte Carlo estimates of *standard errors* ($\times 10^2$) for estimators of the parameter η in the isotropic Winsorized Poisson model

κ_0	η_0	$\hat{\eta}_{\mathcal{P}}$	$\hat{\eta}_{\mathcal{C}}$	$\hat{\eta}_{\mathcal{G}}$
$\log 3$	0.01	2.511	2.391	2.338
$\log 3$	0.02	2.440	2.322	2.269
$\log 3$	0.03	2.325	2.214	2.159
$\log 3$	0.04	2.161	2.053	1.993
$\log 3$	0.05	1.921	1.826	1.760

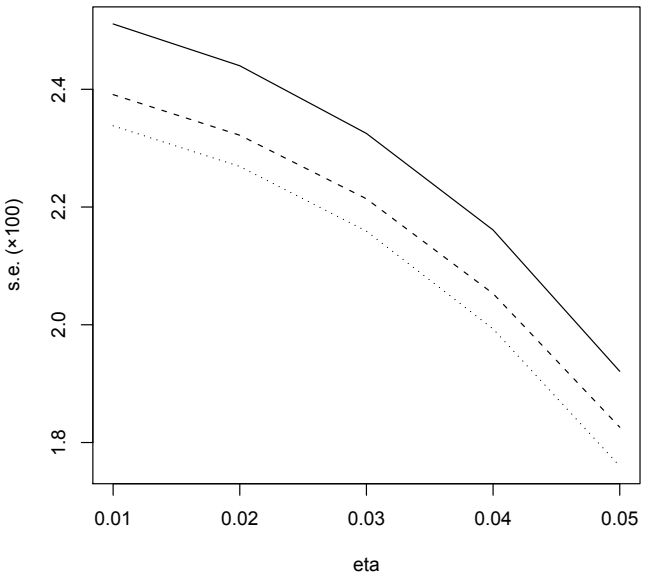

Figure 4.4 Monte Carlo estimates of *standard errors* (s.e.) for estimators of the parameter η in the isotropic Winsorized Poisson model. The solid curve corresponds to Besag's MPLE ($\hat{\eta}_{\mathcal{P}}$), the dotted curve corresponds to Huang and Ogata's MGPLE ($\hat{\eta}_{\mathcal{G}}$), and the dashed curve corresponds to the MCPPLE ($\hat{\eta}_{\mathcal{C}}$).

Table 4.5 Monte Carlo estimates of *bias* ($\times 10^2$) for estimators of the parameter κ in the isotropic Winsorized Poisson model

κ_0	η_0	$\hat{\kappa}_{\mathcal{P}}$	$\hat{\kappa}_{\mathcal{C}}$	$\hat{\kappa}_{\mathcal{G}}$
$\log 3$	0.01	-0.498	-0.509	-0.473
$\log 3$	0.02	-0.515	-0.525	-0.493
$\log 3$	0.03	-0.552	-0.506	-0.447
$\log 3$	0.04	-0.636	-0.546	-0.476
$\log 3$	0.05	-0.563	-0.435	-0.304

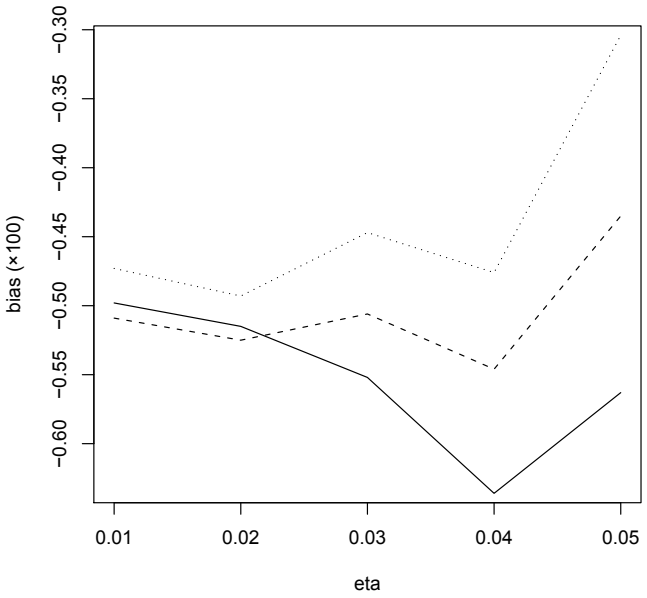

Figure 4.5 Monte Carlo estimates of *bias* for estimators of the parameter κ in the isotropic Winsorized Poisson model. The solid curve corresponds to Besag's MPLE ($\hat{\kappa}_{\mathcal{P}}$), the dotted curve corresponds to Huang and Ogata's MGPLE ($\hat{\kappa}_{\mathcal{G}}$), and the dashed curve corresponds to the MCPPLE ($\hat{\kappa}_{\mathcal{C}}$).

Table 4.6 Monte Carlo estimates of *bias* ($\times 10^2$) for estimators of the parameter η in the isotropic Winsorized Poisson model

κ_0	η_0	$\hat{\eta}_{\mathcal{P}}$	$\hat{\eta}_{\mathcal{C}}$	$\hat{\eta}_{\mathcal{G}}$
$\log 3$	0.01	-0.412	-0.459	-0.452
$\log 3$	0.02	-0.462	-0.550	-0.556
$\log 3$	0.03	-0.523	-0.647	-0.664
$\log 3$	0.04	-0.551	-0.702	-0.734
$\log 3$	0.05	-0.501	-0.672	-0.718

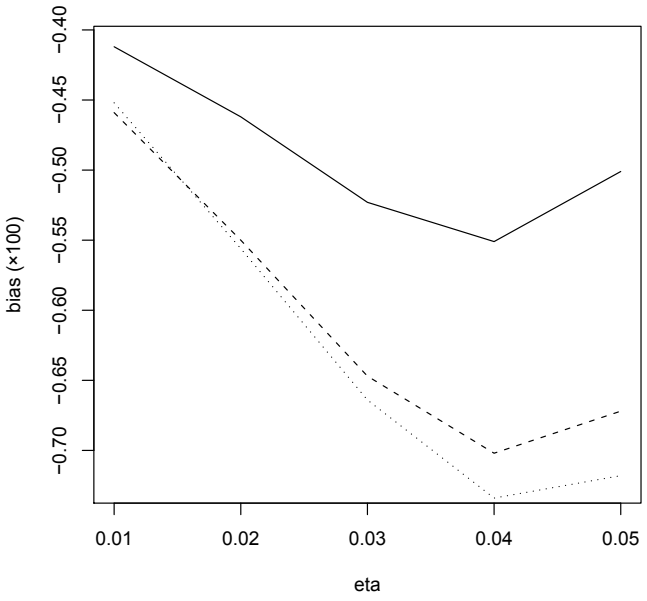

Figure 4.6 Monte Carlo estimates of *bias* for estimators of the parameter η in the isotropic Winsorized Poisson model. The solid curve corresponds to Besag's MPLE ($\hat{\eta}_{\mathcal{P}}$), the dotted curve corresponds to Huang and Ogata's MGPLE ($\hat{\eta}_{\mathcal{G}}$), and the dashed curve corresponds to the MCPPLE ($\hat{\eta}_{\mathcal{C}}$).

Table 4.7 Monte Carlo estimates of *mean square error of the prediction* for estimators in the isotropic Winsorized Poisson model

κ_0	η	$\hat{\boldsymbol{\theta}}_{\mathcal{P}}$	$\hat{\boldsymbol{\theta}}_{\mathcal{C}}$	$\hat{\boldsymbol{\theta}}_{\mathcal{G}}$
$\log 3$	0.01	2.920 (2.913, 2.927)	2.922 (2.915, 2.930)	2.921 (2.914, 2.929)
$\log 3$	0.02	2.936 (2.929, 2.944)	2.939 (2.931, 2.946)	2.938 (2.930, 2.945)
$\log 3$	0.03	2.963 (2.955, 2.971)	2.966 (2.958, 2.973)	2.965 (2.957, 2.972)
$\log 3$	0.04	2.994 (2.986, 3.002)	2.996 (2.988, 3.004)	2.996 (2.998, 3.004)
$\log 3$	0.05	3.087 (3.078, 3.095)	3.090 (3.081, 3.098)	3.090 (3.081, 3.098)

NOTE: Values in parentheses are the bounds of an approximate 90% confidence interval.

is the estimate of $E_{\boldsymbol{\theta}}(Z(s_i)|\boldsymbol{z}(N_i))$, the conditional expectation of the Winsorized Poisson random variable $Z(s_i)$ given values $\boldsymbol{z}(N_i) \equiv \{z(s_j) : s_j \in N_i\}$. Here, $R = 12$ and $\hat{\lambda}_i = \exp\{\hat{\mathcal{A}}_i(z(N_i))\}$ for $\hat{\mathcal{A}}_i(z(N_i)) = \hat{\kappa} + \sum_{s_j \in N_i} \hat{\eta}\,\{z(s_j) - \exp(\hat{\kappa})\}$, where $\hat{\kappa}$ and $\hat{\eta}$ are, respectively, estimates of κ and η based on the realization $\boldsymbol{z}$. Quantity (4.15) was computed for each of the three estimators $\hat{\boldsymbol{\theta}}_{\mathcal{P}} \equiv (\hat{\kappa}_{\mathcal{P}}, \hat{\eta}_{\mathcal{P}})^T$, $\hat{\boldsymbol{\theta}}_C \equiv (\hat{\kappa}_{\mathcal{C}}, \hat{\eta}_{\mathcal{C}})^T$, and $\hat{\boldsymbol{\theta}}_G \equiv (\hat{\kappa}_{\mathcal{G}}, \hat{\eta}_{\mathcal{G}})^T$, and averaged over the $10,000$ realizations in each case. We call each of these averages the Monte Carlo estimates of the *mean square error of the prediction* (m.s.e.p.). Point estimates and approximate confidence intervals for the m.s.e.p. of the three estimators are given in Table 4.7. Note that there appears to be a difference detectable only to the third decimal place between the point estimates of m.s.e.p. for the three estimators, in each parameter value $\boldsymbol{\theta}_0 = (\kappa_0, \eta_0)^T$. However, Table 4.7 shows also that the ranges of values covered by the 90% confidence interval for the m.s.e.p. of any estimator is almost totally overlapped with the range of values covered by the confidence intervals for the m.s.e.p. of the other two. Thus we may conclude that in this case there are not significant differences between the three estimators for any of the considered parameter values $\boldsymbol{\theta}_0 = (\kappa_0, \eta_0)^T$.

We also calculated the asymptotic covariance matrix for each of the estimators of the parameter $\boldsymbol{\theta} = (\kappa, \eta)^T$ by the MCMC method. Define $\theta_1 \equiv \kappa$, $\theta_2 \equiv \eta$, and take $\boldsymbol{\theta} = (\theta_1, \theta_2)^T$. Let $f(\boldsymbol{\theta}; \boldsymbol{z})$ given as in (4.10) represent any of the functions log PL, log GPL, or log CPPL. Then, the surmised asymptotic covariance matrix of the maximizer with respect to $\boldsymbol{\theta}$ of $f(\boldsymbol{\theta}; \boldsymbol{z})$

is the inverse of the Godambe information (4.7),

$$\mathcal{H}(\boldsymbol{\theta})^{-1}\mathcal{J}(\boldsymbol{\theta})\mathcal{H}(\boldsymbol{\theta})^{-1}, \tag{4.17}$$

where the element i, j $(i, j = 1, 2)$ of matrices $\mathcal{H}(\boldsymbol{\theta})$ and $\mathcal{J}(\boldsymbol{\theta})$ is, respectively,

$$\{\mathcal{H}(\boldsymbol{\theta})\}_{i,j} = \mathrm{E}_{\boldsymbol{\theta}}\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j}f(\boldsymbol{\theta};\boldsymbol{Z})\right) \tag{4.18}$$

and

$$\{\mathcal{J}(\boldsymbol{\theta})\}_{i,j} = \mathrm{E}_{\boldsymbol{\theta}}\left(\frac{\partial}{\partial\theta_i}f(\boldsymbol{\theta};\boldsymbol{Z})\frac{\partial}{\partial\theta_j}f(\boldsymbol{\theta};\boldsymbol{Z})\right) \tag{4.19}$$

Here, we generated 15,000 realizations $\boldsymbol{z}$ of the process induced by (4.14) for each of the parameter values $\boldsymbol{\theta}_0 = (\kappa_0, \eta_0)^T$, $\kappa = \log 3$, $\eta_0 = 0.01, 0.02, 0.03, 0.04, 0.05$. For each of these realizations, we computed the differentials $\frac{\partial}{\partial\theta_i}f(\boldsymbol{\theta};\boldsymbol{z})$ and $\frac{\partial^2}{\partial\theta_i\partial\theta_j}f(\boldsymbol{\theta};\boldsymbol{z})$ as given in (4.11) and (4.12), respectively. Then, expectations (4.18) and (4.19) were estimated by the corresponding Monte Carlo averages. Tables 4.8 and 4.9 give estimates of asymptotic standard errors for estimators of κ and η, respectively, and Figures 4.7 and 4.8 show, respectively, the same results graphically. From Figure 4.7 and Table 4.8 we can see that there are just slight differences in asymptotic standard errors for the three estimators of the parameter κ. From Figure 4.7 note that for the dependence parameter values $\eta_0 = 0.01, 0.02$, and 0.03, the MCPPLE of κ $(\hat{\kappa}_{\mathcal{C}})$ appears to have slightly larger values of asymptotic standard errors than the other two but when η_0 goes from 0.04 to 0.05, its values of asymptotic standard errors are slightly smaller than the ones for the MPLE $(\hat{\kappa}_{\mathcal{P}})$ and about the same as the values for the MGPLE $(\hat{\kappa}_{\mathcal{G}})$ which have the largest and smallest values, respectively, of asymptotic standard error for these values of η_0. On the other hand, we find a more defined pattern for asymptotic standard errors of the estimators of η as shown by Figure 4.8. Note that $\hat{\eta}_{\mathcal{P}}$ and $\hat{\kappa}_{\mathcal{G}}$ are the worst and the best behaved, respectively, of the three estimators. Note also that $\hat{\eta}_{\mathcal{C}}$ has significantly better performance than $\hat{\eta}_{\mathcal{P}}$ and close to the performance of $\hat{\eta}_{\mathcal{G}}$. Figure 4.8 shows also that as η_0 goes from 0.01 to 0.05, the difference in the values of asymptotic standard error between $\hat{\eta}_{\mathcal{P}}$ and any of $\hat{\eta}_{\mathcal{C}}$ and $\hat{\eta}_{\mathcal{G}}$ increases with the largest difference taking place at $\eta_0 = 0.05$. In contrast, note that there appears that the difference in values of asymptotic standard errors of $\hat{\eta}_{\mathcal{C}}$ and $\hat{\eta}_{\mathcal{G}}$ keep small and roughly constant as η_0 increases from 0.01 to 0.05.

Table 4.8 Monte Carlo estimates of *asymptotic standard errors* ($\times 10^2$) for estimators of the parameter κ in the isotropic Winsorized Poisson model

κ_0	η_0	$\hat{\kappa}_{\mathcal{P}}$	$\hat{\kappa}_{\mathcal{C}}$	$\hat{\kappa}_{\mathcal{G}}$
$\log 3$	0.01	6.110	6.182	6.117
$\log 3$	0.02	6.532	6.590	6.532
$\log 3$	0.03	7.109	7.144	7.096
$\log 3$	0.04	7.967	7.947	7.917
$\log 3$	0.05	9.514	9.351	9.333

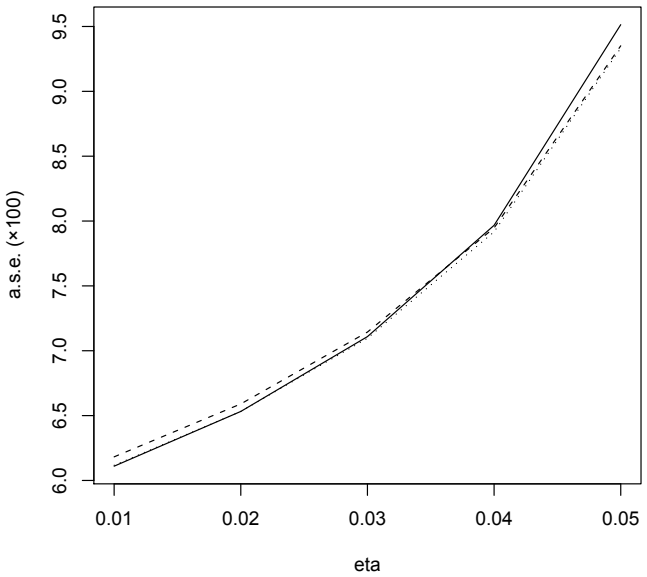

Figure 4.7 Monte Carlo estimates of *asymptotic standard errors* (a.s.e.) for estimators of the parameter κ in the isotropic Winsorized Poisson model. The solid curve corresponds to Besag's MPLE ($\hat{\kappa}_{\mathcal{P}}$), the dotted curve corresponds to Huang and Ogata's MG-PLE ($\hat{\kappa}_{\mathcal{G}}$), and the dashed curve corresponds to the MCPPLE ($\hat{\kappa}_{\mathcal{C}}$).

Table 4.9 Monte Carlo estimates of *asymptotic standard errors* ($\times 10^2$) for estimators of the parameter η in the isotropic Winsorized Poisson model

κ_0	η_0	$\hat{\eta}_{\mathcal{P}}$	$\hat{\eta}_{\mathcal{C}}$	$\hat{\eta}_{\mathcal{G}}$
$\log 3$	0.01	2.446	2.469	2.448
$\log 3$	0.02	2.365	2.371	2.347
$\log 3$	0.03	2.242	2.220	2.191
$\log 3$	0.04	2.031	1.968	1.932
$\log 3$	0.05	1.676	1.563	1.515

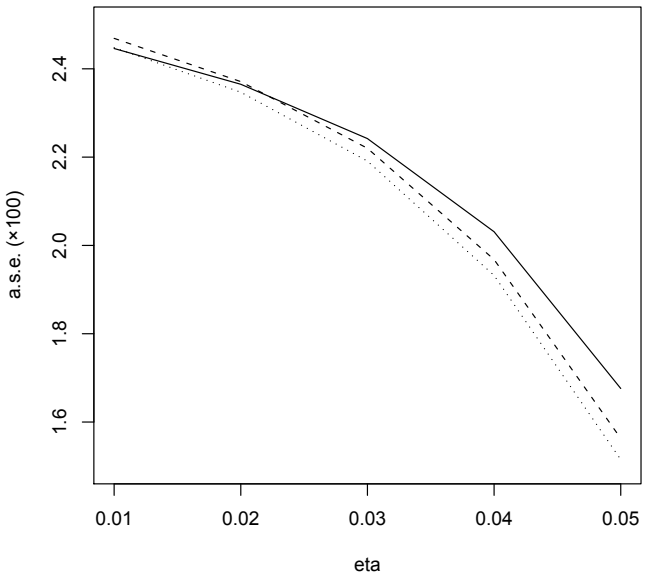

Figure 4.8 Monte Carlo estimates of *asymptotic standard errors* (a.s.e.) for estimators of the parameter η in the isotropic Winsorized Poisson model. The solid curve corresponds to Besag's MPLE ($\hat{\eta}_{\mathcal{P}}$), the dotted curve corresponds to Huang and Ogata's MG-PLE ($\hat{\eta}_{\mathcal{G}}$), and the dashed curve corresponds to the MCPPLE ($\hat{\eta}_{\mathcal{C}}$).

As we have seen, Huang and Ogata's MGPLE, in the way we have defined it here, showed significantly better performance than Besag's MPLE. But it was just slightly better than the MCPPLE. On the other hand, the calculation of the MGPLE is much more difficult to implement than the MPLE and the MCPPLE. Essentially, the difficulty to compute the MGPLE comes from the complexity to calculate the normalizing terms

$$\sum_{\boldsymbol{t}\in\Omega_k} \exp\{Q(\boldsymbol{t}, \boldsymbol{z}^{g_k}; \boldsymbol{\theta})\} \tag{4.20}$$

appearing in (4.10), when $|g_k|$ gets large. Note that each of these terms are the analog of the partition function (3.2) of the joint pmf (3.1) for the conditional pmfs $p_{\boldsymbol{\theta}}(\boldsymbol{z}_{g_k}|\boldsymbol{z}^{g_k})$. For example, when the group g_k defined for site s_k consists of 5 sites (i.e, s_k is in the interior of the four-nearest neighbor regular lattice), then (4.20) is a sum of 13^5 (=371293) exponentials (for $R = 12$). Add to this the fact that the terms (4.20) have to be updated in every step of the iterative maximization algorithm to calculate the objective function (4.10) and the first and second order differentials (4.11) and (4.12). Thus an enormous computational effort is required to calculate the MGPLE for groups of size 5, even for small data sets, as for the 10×10 square lattice used in our simulation studies. In contrast, the calculation of the MCPPLE was nearly as easy as the MPLE, and its performance was significantly better than the MPLE and very close to the MGPLE. Therefore, we believe that the MCPPLE would be a more viable alternative of estimation than the MPLE and the MGPLE for large data sets, and for situations where the truncation parameter R of the Winsorized model under consideration is large.

CHAPTER 5. MODELS WITH DIRECTIONAL DEPENDENCE

A second simulation study was implemented to study the finite-sample properties of the MCPPLE under a directional spatial dependence structure on the regular square lattice $\mathbb{Z}^2_{\mathcal{L}\times\mathcal{L}} \equiv \{(u_i, v_i) : (u_i, v_i) \in \{1, 2, \ldots, \mathcal{L}\} \times \{1, 2, \ldots, \mathcal{L}\}\}$ with a four-nearest neighbor configuration for the neighborhoods N_i, $i = 1, \ldots, n$; $n \equiv |\mathbb{Z}^2_{\mathcal{L}\times\mathcal{L}}|$. We assume free boundary conditions for the square lattice; that is, $Z((u_i, v_i)) = 0$ whenever $(u_i, v_i) \notin \mathbb{Z}^2_{\mathcal{L}\times\mathcal{L}}$. Let $s_i \equiv (u_i, v_i)$ were u_i denotes the integer value for horizontal grid location and v_i the corresponding integer for vertical grid location. Also, define $N_{1i} \equiv \{(u_i - 1, v_i), (u_i + 1, v_i)\}$ and $N_{2i} \equiv \{(u_i, v_i - 1), (u_i, v_i + 1)\}$; $N_i = N_{1i} \cup N_{2i}$, $i = 1, \ldots, n$. We considered an anisotropic (directional dependence) model by modeling the set of dependencies $\{\eta_{ij} : i, j = 1, \ldots, n\}$ of the model induced by (4.9) as

$$\eta_{ij} = \begin{cases} \eta_1 & \text{if } s_j = (u_i - 1, v_i) \\ \eta_1 & \text{if } s_j = (u_i + 1, v_i) \\ \eta_2 & \text{if } s_j = (u_i, v_i - 1) \\ \eta_2 & \text{if } s_j = (u_i, v_i + 1) \\ 0 & \text{otherwise,} \end{cases}$$

i.e., $\eta_{ij} = \eta_1$ if $s_j \in N_{1i}$, $\eta_{ij} = \eta_2$ if $s_j \in N_{2i}$, and $\eta_{ij} = 0$ otherwise. In addition, we make $\kappa_i = \kappa, i = 1, \ldots, n$. Then, under this dependence structure, the negpotential function (4.9) becomes

$$Q(\boldsymbol{z}) = \sum_{1\le i\le n} \Big[\big(\kappa - \eta_1 \exp(\kappa)|N_{1i}| - \eta_2 \exp(\kappa)|N_{2i}|\big) z(s_i) - \log\{z(s_i)!\}\Big] + \eta_1 \sum_{\{s_i,s_j\}\in\mathcal{C}_1} z(s_i)z(s_j) + \eta_2 \sum_{\{s_i,s_j\}\in\mathcal{C}_2} z(s_i)z(s_j), \quad (5.1)$$

where $|N_{1i}|$ denotes the cardinality of N_{1i}, $|N_{2i}|$ denotes the cardinality of N_{2i}, $\mathcal{C}_1 \equiv \{\{s_i, s_j\} : s_j \in N_{1i}\}$, and $\mathcal{C}_2 \equiv \{\{s_i, s_j\} : s_j \in N_{2i}\}$; $\mathcal{C} = \mathcal{C}_1 \cup \mathcal{C}_2$.

We used the Gibbs sampler (Geman and Geman 1984), roughly described in Chapter 4, to generate the process induced by the negpotential function (5.1) defined over the 10×10 square lattice $\mathbb{Z}^2_{10\times 10} \equiv \{(u_i, v_i) : (u_i, v_i) \in \{1, 2, \ldots, 10\} \times \{1, 2, \ldots, 10\}\}$. In our simulation, the parameter η_1 assumes values $\eta_{1_0} = 0.03, 0.04$, and 0.05, the parameter η_2 assumes values $\eta_{2_0} = 0.00, 0.01$, and 0.02, and, for all combinations of these, the true parameter κ value is set to be always $\kappa_0 = \log 3$ as in the first study. All η_1, η_2, and κ are assumed unknown when these are estimated. The truncation parameter R of the local conditional distributions $p(z(s_i)|z(N_i))$ in (4.1)) was set fixed and known to $R = 12$ in all cases. Convergence of the iterative simulation was monitored using the method of multiple sequences proposed by Gelman and Rubin (1992). Convergence to the target distribution (i.e., the joint pmf induced by (5.1)) was fast for all parameter values and the burn-in period for the chain was set at 400 iterations in all cases. After the burn-in period, the chain was subsampled every 10 iterations to reduce autocorrelation. For each combination of true parameter values η_{1_0}, η_{2_0}, and κ_0, we generated $20,000$ realizations of the process, and from each of these we calculated the MPLE and the MCPPLE of the parameter $\boldsymbol{\theta} \equiv (\kappa, \eta_1, \eta_2)^T$ (i.e., we obtained 20,000 realizations of each estimator). Because of the additional calculation complexity added by considering the estimation of one more parameter, we ran out of computational resources to obtain the MGPLE. Hereafter, $\hat{\eta}_{1\mathcal{P}}$ and $\hat{\eta}_{1\mathcal{C}}$ denote, respectively, the MPLE and the MCPPLE of the parameter η_1, $\hat{\eta}_{2\mathcal{P}}$ and $\hat{\eta}_{2\mathcal{C}}$ denote, respectively, the MPLE and the MCPPLE of the parameter η_2, and $\hat{\kappa}_{\mathcal{P}}$ and $\hat{\kappa}_{\mathcal{C}}$ represent the corresponding estimates for the parameter κ.

In this simulation study we compare the performance of the MCPPLE to the MPLE respect to *mean square error*: $E_{\boldsymbol{\theta_0}}(\hat{\theta} - \theta_0)^2$, *bias*: $E_{\boldsymbol{\theta}_0}(\hat{\theta} - \theta_0)$, and *standard error*: $\sqrt{E_{\boldsymbol{\theta}_0}(\hat{\theta} - E_{\boldsymbol{\theta}_0}(\hat{\theta}))^2}$; where, $\boldsymbol{\theta}_0 \equiv (\kappa_0, \eta_{1_0}, \eta_{2_0})^T$ represents the true value of the parameter vector $\boldsymbol{\theta} \equiv (\kappa, \eta_1, \eta_2)^T$, $E_{\boldsymbol{\theta_0}}$ denotes expectation under $\boldsymbol{\theta}_0$, the scalar θ_0 denotes one of κ_0, η_{1_0}, and η_{2_0}, and $\hat{\theta}$ represents the estimator of θ_0. For $M = 20,000$, these were estimated, respectively, by $\sum_{i=1}^{M}(\hat{\theta}_i - \theta_0)^2/M$, $\sum_{i=1}^{M}(\hat{\theta}_i - \theta_0)/M$, and $\sqrt{\sum_{i=1}^{M}(\hat{\theta}_i - \bar{\hat{\theta}})^2/(M-1)}$; $\bar{\hat{\theta}} = \sum_{i=1}^{M}\hat{\theta}_i/M$. We also computed

Table 5.1 Monte Carlo estimates of *mean square errors* ($\times 10^2$) for estimators of the parameter κ in the anisotropic Winsorized Poisson model

κ_0	η_{1_0}	η_{2_0}	$\hat{\kappa}_{\mathcal{P}}$	$\hat{\kappa}_{\mathcal{C}}$
$\log 3$	0.05	0.00	0.508 (0.496, 0.521)	0.470 (0.461, 0.480)
$\log 3$	0.04	0.01	0.519 (0.503, 0.535)	0.468 (0.458, 0.477)
$\log 3$	0.03	0.02	0.499 (0.485, 0.513)	0.460 (0.450, 0.469)

NOTE: Values in parentheses are the bounds of an approximate 90% confidence interval.

Table 5.2 Monte Carlo estimates of *mean square errors* ($\times 10^2$) for estimators of the parameter η_1 in the anisotropic Winsorized Poisson model

κ_0	η_{1_0}	η_{2_0}	$\hat{\eta}_{1\mathcal{P}}$	$\hat{\eta}_{1\mathcal{C}}$
$\log 3$	0.05	0.00	0.107 (0.106, 0.109)	0.100 (0.098, 0.102)
$\log 3$	0.04	0.01	0.113 (0.111, 0.115)	0.105 (0.103, 0.107)
$\log 3$	0.03	0.02	0.116 (0.114, 0.118)	0.108 (0.106, 0.111)

NOTE: Values in parentheses are the bounds of an approximate 90% confidence interval.

approximate confidence intervals for the mean square error of the estimator $\hat{\theta}$ by using the central limit theorem, taking $\sqrt{\sum_{j=1}^{M}[(\hat{\theta}_j - \theta_0)^2 - \sum_{i=1}^{M}(\hat{\theta}_i - \theta_0)^2/M]^2/(M-1)}/\sqrt{M}$ as the standard error of $\sum_{i=1}^{M}(\hat{\theta}_i - \theta_0)^2/M$.

Tables 5.1, 5.2, and 5.3 give point estimates and 90% confidence intervals of mean square errors for the MPLE and MCPPLE of the parameters κ, η_1, and η_2, respectively. Point estimates and 90% confidence intervals indicate that in all three cases the MPLE has larger mean square errors than the MCPPLE for any combination of true parameter values κ_0, η_{1_0}, and η_{2_0}. Note also that there is not overlapping of the 90% confidence intervals for the mean square error of the MPLE and the MCPPLE in any of the three cases, for any combination of true parameter values κ_0, η_{1_0}, and η_{2_0}.

Tables 5.4, 5.5, and 5.6 give estimates of biases and standard errors for the MPLE and

Table 5.3 Monte Carlo estimates of *mean square errors* ($\times 10^2$) for estimators of the parameter η_2 in the anisotropic Winsorized Poisson model

κ_0	η_{1_0}	η_{2_0}	$\hat{\eta}_{2\mathcal{P}}$	$\hat{\eta}_{2\mathcal{C}}$
$\log 3$	0.05	0.00	0.114 (0.112, 0.116)	0.104 (0.102, 0.105)
$\log 3$	0.04	0.01	0.118 (0.116, 0.120)	0.108 (0.106, 0.109)
$\log 3$	0.03	0.02	0.118 (0.116, 0.120)	0.109 (0.107, 0.111)

NOTE: Values in parentheses are the bounds of an approximate 90% confidence interval.

Table 5.4 Monte Carlo estimates of *standard errors* ($\times 10^2$) and *biases* ($\times 10^2$) for estimators of the parameter κ in the anisotropic Winsorized Poisson model

			$\hat{\kappa}_{\mathcal{P}}$		$\hat{\kappa}_{\mathcal{C}}$	
κ_0	η_{1_0}	η_{2_0}	s.d.	bias	s.d.	bias
$\log 3$	0.05	0.00	7.068	-0.936	6.801	-0.889
$\log 3$	0.04	0.01	7.148	-0.898	6.784	-0.857
$\log 3$	0.03	0.02	7.009	-0.894	6.726	-0.854

MCPPLE of the parameters κ, η_1, and η_2, respectively. Note that the estimates of standard errors follow the same pattern as the mean square errors for the MPLE and MCPPLE in all three cases, i.e., the MPLE has larger standard errors than the MCPPLE for any combination of true parameter values κ_0, η_{1_0}, and η_{2_0}. Table 5.4 shows also that for estimators of the parameter k, the MCPPLE ($\hat{\kappa}_{\mathcal{C}}$) has smaller bias than the MPLE ($\hat{\kappa}_{\mathcal{P}}$) for any combination of κ_0, η_{1_0}, and η_{2_0}. On the other hand, Tables 5.5, and 5.6 show that the MCPPLE of η_1 ($\hat{\eta}_{1\mathcal{C}}$) and the MCPPLE of η_2 ($\hat{\eta}_{2\mathcal{C}}$) have larger bias than the MPLE of η_1 ($\hat{\eta}_{1\mathcal{P}}$) and the MPLE of η_2 ($\hat{\eta}_{2\mathcal{P}}$), respectively.

Table 5.5 Monte Carlo estimates of *standard errors* ($\times 10^2$) and *biases* ($\times 10^2$) for estimators of the parameter η_1 in the anisotropic Winsorized Poisson model

			$\hat{\eta}_{1\mathcal{P}}$		$\hat{\eta}_{1\mathcal{C}}$	
κ_0	η_{1_0}	η_{2_0}	s.d.	bias	s.d.	bias
$\log 3$	0.05	0.00	3.219	-0.615	3.056	-0.808
$\log 3$	0.04	0.01	3.318	-0.576	3.160	-0.739
$\log 3$	0.03	0.02	3.368	-0.549	3.218	-0.677

Table 5.6 Monte Carlo estimates of *standard errors* ($\times 10^2$) and *biases* ($\times 10^2$) for estimators of the parameter η_2 in the anisotropic Winsorized Poisson model

			$\hat{\eta}_{2\mathcal{P}}$		$\hat{\eta}_{2\mathcal{C}}$	
κ_0	η_{1_0}	η_{2_0}	s.d.	bias	s.d.	bias
$\log 3$	0.05	0.00	3.343	-0.448	3.182	-0.478
$\log 3$	0.04	0.01	3.392	-0.501	3.233	-0.562
$\log 3$	0.03	0.02	3.398	-0.537	3.239	-0.630

CHAPTER 6. CONCLUSION

Whereas Huang and Ogata's approach of the maximum generalized pseudo-likelihood (MGPL) method to improve the performance of Besag's maximum pseudo-likelihood (MPL) may work well for Markov random fields where the site variables have auto-normal structure or for variables that assume only two values, generally speaking, this approach may become very difficult to implement computationally, as we have seen for the Winsorized Poisson auto-models in the past chapters of this dissertation. For models where each site variable, conditional on its neighbors, follows the distribution of a Winsorized Poisson random variable, the MGPL principle of defining groups of adjacent sites for each site in the lattice to compute the GPL function and increase the size of these groups until achieving a significantly better performance of the corresponding estimator (i.e., the MGPLE) than the MPL estimator (i.e., the MPLE) may be marred by the same problem of awkward normalizing factors as we have seen for the implementation of the full maximum likelihood method. In contrast, our proposed maximum conditional pairwise pseudo-likelihood (MCPPL) method, described fully in the past chapters, is as easy to implement as the MPL method, and the corresponding estimator (i.e., the MCPPLE) has significantly better performance than the MPLE. Thus we suggest in this dissertation that for the case of Winsorized Poisson auto-models, the MCPPL method is a more viable alternative of estimation than the MGPL and MPL methods.

As regards of future research, extensions of the MCPPL method to spatial processes other than the Winsorized Poisson auto-models considered in this dissertation would be a primary goal. Illustration of the MCPPL method with real data sets would be also a pertinent goal for future work.

BIBLIOGRAPHY

Arnold, B. C., and Strauss, D. (1994), "Pseudolikelihood Estimation: Some Examples," *The Indian Journal of Statistics*, Ser. B, 53, 233–243.

Besag, J. E. (1974), "Spatial Interaction and the Statistical Analysis of Lattice Systems" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 36, 192–236.

Besag, J. E. (1975), "Statistical Analysis of Non-Lattice Data," *The Statistician*, 24, 179–195.

Besag, J. E. (1977), "Efficiency of Pseudolikelihood Estimation for Simple Gaussian Fields," *Biometrika*, 64, 616–618.

Besag, J. E. (1986), "On the Statistical Analysis of Dirty Pictures" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 48, 259–302.

Casella, G., and Berger, R. L. (2002), *Statistical Inference*, Pacific Grove: Duxbury.

Chow, Y. S., and Teicher, H. (1988), *Probability Theory: Independence, Interchangeability, Martingales*, New York: Springer.

Clifford, P. (1990), "Markov Random Fields in Statistics," in *Disorder in Physical Systems*, ed. G.R. Grimmett and D.J.A. Welsh, Oxford University Press, 19–32.

Comets, F. (1992), "On Consistency of a Class of Estimators for Exponential Families of Markov Random Fields on the Lattice," *The Annals of Statistics*, 20, 455–468.

Comets, F. (1994), "Large Deviations and Consistent Estimates for Gibbs Random Fields," in *Proceedings, 1994 IEEE-IMS Workshop on Information Theory and Statistics*, Alejandria, Virginia, 43–.

Comets, F. (1997), "Detecting Phase Transition for Gibbs Measures," *The Annals of Applied Probability*, 7, 545–563.

Cox, D. R., and Reid, N. (2004), "A Note on Pseudolikelihood Constructed from Marginal Densities," *Biometrika*, 91, 729–737.

Cressie, N. A. C. (1993), *Statistics for Spatial Data*, New York: Wiley.

Cressie, N., and Lele, S. (1992), "New Models for Markov Random Fields," *Journal of Applied Probabilty*, 29, 877–884.

Crowder, M. J. (1986), "On Consistency and Inconsistency of Estimating Equations," *Econometric Theory*, 2, 305–330.

Föllmer, H., and Orey, S. (1988), "Large Deviation for the Empirical Field of a Gibbs Measure," *The Annals of Probability*, 16, 961–977.

Geman, S., and Geman, D. (1984), "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721–741.

Geman, S., and Graffigne, C. (1986), "Markov Random Field Image Models and Their Applications to Computer Vision," in *Proceedings of the International Congress of Mathematicians*, Berkeley, California, 1496–1517.

Gelman, A., and Rubin, D. B. (1992), "Inference from Iterative Simulation Using Multiple Sequences," *Statistical Science*, 7, 457–511.

Geyer, C.J. (1991), "Markov Chain Monte Carlo Maximum Likelihood," in *Computing Science and Statistics: 23rd Symposium on the Interface*, ed. E. M. Keramidas, Fairfax Station, VA: Interface Foundation of North America, pp. 156–163.

Geyer, C.J., and Thompson, E. A. (1992), "Constrained Monte Carlo Maximum Likelihood for Dependent Data," *Journal of the Royal Statistical Society*, Ser. B, 54, 657–699.

Gidas, B. (1986), "Consistency of Maximum Likelihood and Pseudolikelihood Estimators for Gibbs Distributions," in *Stochastic Differential Systems with Applications to Electronic/Computer Engineering, Control Theory and Operations Research* (W. Fleming and P. L. Lions, eds.) 1–17. Springer, Berlin.

Gidas, B. (1993), "Parameter Estimation for Gibbs Distributions from Fully Observed Data," in *Markov Random Fields: Theory and Applications* (R. Chellappa and A. Jain, eds.) 471–498. Academic Press, New York.

Godambe, V. P. (1960), "An Optimum Property of Regular Maximum Likelihood Estimation," *The Annals of Mathematical Statistics*, 31, 1208–1211.

Goulard, M., Särkkä, A., and Grabarnik, P. (1996), "Parameter Estimation for Marked Gibbs Point Processes Through the Maximum Pseudolikelihood Method," *Scandinavian Journal of Statistics*, 23, 365–379.

Guyon, X. (1995), *Random Fields on a Network: Modeling, Statistics and Applications*, New York: Springer.

Guyon, X., and Künsch, H. R. (1992), "Asymptotic Comparison of Estimators in the Ising Model," in P. Barone, A. Frigessi and M. Piccioni (eds), Stochastic Models, Statistical Methods, and Algorithms in Image Analysis, Lecture Notes in Statistics 74, Springer, Berlin.

Hammersley, J. M., and Clifford, P. (1971), "Markov Fields on Finite Graphs and Lattices," unpublished manuscript, Oxford University.

Hastings, W. K. (1970), "Monte Carlo Sampling Methods Using Markov Chains and Their Applications," *Biometrika*, 57, 97–109.

Heagerty, P. J., and Lele, S. R. (1998), "A Composite Likelihood Approach to Binary Spatial Data," *Journal of The American Statistical Association*, 93, 1099–1111.

Hjort, N. L., and Omre, H. (1994), "Topics in Spatial Statistics," *Scandinavian Journal of Statistics*, 21, 289–357.

Huang, F., and Ogata, Y. (1999), "Improvements of the Maximum Pseudo-Likelihood Estimators in Various Spatial Statistical Models," *Journal of Computational and Graphical Statistics*, 8, 510–530.

Huang, F., and Ogata, Y. (2002), "Generalized Pseudo-Likelihood Estimates for Markov Random Fields on Lattice," *Annals of the Institute of Statistical Mathematics*, 54, 1–18.

Jensen, J. L., and Künsch, H. R. (1994), "On asymptotic Normality of Pseudolikelihood Estimates for Pairwise Interaction Processes," *Annals of the Institute of Statistical Mathematics*, 46, 475–486.

Jensen, J. L., and Møller, J. (1991), "Pseudolikelihood for Exponential Family Models of Spatial Point Processes," *The Annals of Applied Probability*, 1, 445–461.

Kaiser, M. S. (2007), "Statistical Dependence in Markov Random Field Models," Preprint 2007, Department of Statistics, Iowa State University, Ames, Iowa.

Kaiser, M. S., and Cressie, N. (1997), "Modeling Poisson Variables With Positive Spatial Dependence," *Statistics & Probability Letters*, 35, 423–432.

Kaiser, M. S., and Cressie, N. (2000), "The Construction of Multivariate Distributions from Markov Random Fields," *Journal of Multivariate Analysis*, 73, 199–220.

Kaiser, M. S., Cressie, N., and Lee, J. (2002), "Spatial Mixture Models based on Exponential Family Conditional Distributions," *Statistica Sinica*, 12, 449–474.

Kaiser, M. S., Caragea, P. C., and Furukawa, K. (2004), "Centering the Effects of Neighbors in Markov Random Field Models," The Graybill Conference: Spatial Statistics Agricultural, Ecological, and Environmental Applications, Colorado State University. Invited talk.

Kuk, A. Y. C., and Nott, D. J. (2000), "A Pairwise Likelihood Approach to Analyzing Correlated Binary Data," *Statistics & Probability Letters*, 47, 329–335.

Kullback, S., and Leibler, R. A. (1951), "On Information and Sufficiency," *The Annals of Mathematical Statistics*, 22, 79–86.

Lehmann, E., and Romano, J. P. (2004), *Testing Statistical Hypotheses*, New York: Springer.

Liang, K. Y., and Zeger, S. L. (1986), "Longitudinal Data Analysis Using Generalized Linear Models," *Biometrika*, 73, 13–22.

Lindsay, B. G. (1988), "Composite Likelihood Methods," *Contemporary Mathematics*, 80, 221–239.

Mase, S. (1995), "Consistency of the Maximum Pseudo-Likelihood Estimator of Continuous State Space Gibbsian Processes," *The Annals of Applied Probability*, 5, 603–612.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953), "Equations of State Calculations by Fast Computing Machines," *Journal of Chemical Physics*, 21, 1087–1091.

Moyeed, R.A., and Baddeley, A.J. (1991), "Stochastic Approximation of the MLE for a Spatial Point Pattern," *Scandinavian Journal of Statistics*, 18, 39–50.

Nott, D. J., and Rydén, T. (1999), "Pairwise Likelihood Methods for Inference in Image Models," *Biometrika*, 86, 661–676.

Penttinen, A. (1984), "Modelling Interaction in Spatial Point Patterns: Parameter Estimation by the Maximum Likelihood Method," *Jyvaskyla Studies in Computer Science, Economics and Statistics*, 7.

Pickard, D. K. (1987), "Inference for Discrete Markov Fields: The Simplest Nontrivial Case," *Journal of The American Statistical Association*, 82, 90–96.

Song, P. X.-K. (2007), *Correlated Data Analysis: Modeling, Analytics, and Applications*, New York: Springer.

Taylor, A. E., and Mann, W. R. (1983), *Advanced Calculus*, New York: Wiley.

Varin, D., Vidoni, P. (2005), "A Note on Composite Likelihood Inference and Model Selection," *Biometrika*, 92, 519–528.

Varin, D. (2008), "On Composite Marginal Likelihoods," *AStA Advances in Statistical Analysis*, 92, 1–28.

White, H. (1982), "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1–24.

Younes, L. (1988), "Estimation and Annealing for Gibbsian Fields," *Annales de l'Institut Henri Poincaré Probabilités et Statistiques*, 24, 269–294.